Timing and probability of crosstalk in a dense CMOS SPAD array in pulsed TOF applications

SAHBA JAHROMI* AND JUHA KOSTAMOVAARA
University of Oulu, Faculty of Information Technology and Electrical Engineering, Circuits and Systems (CAS) Research Unit Oulu, Finland
*sahba.jahromi@oulu.fi

Abstract: As the distance between neighboring devices in large CMOS single-photon avalanche diode (SPAD) arrays is reduced for improving the density, increased crosstalk becomes an important issue, limiting the maximum practical fill factor of the array. In this study, the temporal correlation of crosstalk events, as well as the crosstalk probability, and their dependence on parameters, such as the illumination wavelength and intensity, and the distance between SPADs, are investigated via measurement of a ~45%-fill factor CMOS SPAD array fabricated using 0.35-µm high-voltage CMOS technology. The SPADs have 24 µm × 24 µm square-shaped active areas, and all devices share a common deep-N-well cathode. On-chip time-to-digital converters with 65-ps resolution are used to measure the timing of crosstalk events in “coincidence measurements.” For the crosstalk measurements, the internal noise in one SPAD is used to produce crosstalk events in the neighboring devices. The measurement results indicate both optical and electrical crosstalk with the crosstalk events, having a specific temporal distribution. The crosstalk probability in the first two adjacent pixels is found to be 0.3% and 0.01%, with a distribution having full widths at half maximum (FWHMs) of 700 and 400 ps, respectively. In pulsed time-of-flight measurements, when one SPAD is triggered with external short-pulsed (FWHM of approximately 200 ps) illumination, extra correlated noise in the adjacent SPADs added to the crosstalk noise, increasing the correlated noise considerably. This additional noise was a secondary effect of the absorbed laser photons deep in the substrate.

© 2018 Optical Society of America under the terms of the OSA Open Access Publishing Agreement

OCIS codes: (230.5160) Photodetectors; (040.6070) Solid state detectors; (140.2020) Diode lasers, (140.3440) Laser-induced breakdown.

References and links
1. Introduction

Single-photon avalanche diode (SPAD) arrays have been established as single-photon detectors of high precision (approximately 50–100 ps) in a variety of applications, such as three-dimensional (3D) imaging, laser ranging, and spectroscopy [1–3]. Complementary metal-oxide semiconductor (CMOS) SPAD arrays are of particular interest owing to the possibility of integration with photon-counting or time-discrimination circuits, giving rise to compact single-chip one/two-dimensional (2D) receivers. In principle, SPAD is a p-n junction that is reverse-biased above its breakdown voltage, so that a single charge carrier injected into the depletion layer can trigger a self-sustaining avalanche. If the primary carrier is photon-generated, the leading edge of the avalanche pulse marks (with picosecond-timing jitter) the arrival time of the detected photon [4]. One of the limiting characteristics of SPAD arrays is correlated noise. The two known sources of correlated noise in Geiger-mode avalanche photodiodes arrays are afterpulsing and crosstalk.

Afterpulsing is caused by traps with energy levels near the edges of the bandgap. During each avalanche, some carriers may become trapped in these energy levels and subsequently released. The released carriers can cause another avalanche, generating afterpulses that are correlated to the previous avalanche of the same SPAD [5]. The release time constant of these traps is on the order of few nanoseconds, with an afterpulsing probability on the order of 0.5–
2\% [6, 7]. In pulsed time-of-flight (TOF) applications, unless the pulsing rate is hundreds of megahertz, afterpulsing is not an important issue, as it can be easily differentiated from a photon-generated avalanche.

Crosstalk pulses are unwanted avalanches in one device (detector of crosstalk noise) that are caused by an avalanche in another device of the array (emitter of crosstalk noise), leading to a false detection. Crosstalk may limit the dynamic range of the SPAD detector array and lead to a blurred image in imaging applications (e.g., in the case of two adjacent SPADs with a considerable difference in incident light intensity, the crosstalk from the neighboring SPAD might dominate the photons coming from the target in the less-illuminated SPAD, leading to a false measurement).

Figure 1 shows different crosstalk mechanisms in two neighboring devices. For simplicity, the devices are reduced to p-n junctions, and the third dimension is not shown. The SPADs share a common terminal, as is the case in high-density arrays [3, 8]. During an avalanche in one SPAD, photons are emitted owing to the electroluminescence effect as a result of the relaxation of hot-carriers generated in response to the passage of a large current in a strong electric field. A portion of these photons travel in direct and indirect optical paths, finding their way to neighboring pixels, leading to optical crosstalk [3, 9]. The crosstalk may also be electrical, i.e., some of the carriers generated during the avalanche may exit the depletion region and diffuse laterally, eventually reaching the depletion region of another SPAD [10, 11]. SPAD arrays, in which each pixel has its own separate terminals, have negligible electrical crosstalk because minority carriers can no longer diffuse to the nearby SPAD. There can also be a combination of both optical and electrical crosstalk, as shown in Fig. 1: a secondary photon generated as a result of an avalanche in the emitter device travels to the vicinity of another device and is absorbed; then, the generated minority carrier might diffuse to the depletion region of the adjacent device.

![Fig. 1. 2D representation of direct (1) and indirect (2) optical crosstalk, as well as electrical crosstalk (3) mechanisms and their combination (4).](image)

The photons or carriers that reach the neighboring SPADs may trigger an avalanche, yielding incorrect measurement results. The occurrence probability of this effect is called the crosstalk probability and is directly dependent on the energy dissipated during an avalanche event in the emitter, as both the electroluminescence intensity and number of carriers diffusing out of the depletion region are related to the number of carriers in an avalanche pulse.

Although studies and experiments have been performed on crosstalk probability and its dependence on various parameters (distance between two devices, excess voltage, die thickness, etc [9, 11–18]), temporal correlation of the crosstalk coincidences in SPAD arrays...
has not been discussed extensively or has been reported with limited timing resolutions [11, 12, 17, 19, 20] (a more detailed review is presented in the Discussion section).

In this paper, the crosstalk and its characteristics in the first two immediate neighbors of an avalanching device are studied, with a focus on the measurement of the temporal behavior of this noise source. The crosstalk is assessed through “coincidence measurements” with a high-dark count rate (DCR) SPAD acting as the crosstalk source. Then, to investigate the correlated noise in a practical pulsed TOF application, coincidence measurements were repeated with an SPAD triggered with pulsed laser illumination instead of own noise.

In the following sections, first, the SPAD array used in the correlated noise measurements is described. Then, the setups and processes for both of the aforementioned measurements are presented, and the methodologies are explained. Finally, the results are discussed, and conclusions are drawn.

2. SPAD array IC

The SPAD array is realized using 0.35-μm high-voltage (HV) CMOS technology. The receiver consists of 81 SPADs in a 9 × 9 square array, and timing of the breakdowns is measured by on-chip time-to-digital converters (TDCs) with 65-ps resolution. The chip includes 10 TDCs, one measures the timing of an electrical start signal marking the laser pulse being transmitted and the other nine, the photon arrival timestamp in nine SPADs. As a result, during each measurement, one 3 × 3 subarray can be enabled to detect photons. The breakdown time from all nine active SPADs are measured simultaneously with respect to the start signal (for details, refer to [21]).

As mentioned previously, in this study, the effect of crosstalk from one SPAD is measured in the next two neighboring pixels. Figure 2(a) shows the cross-section of these three adjacent SPADs with their respective dimensions. The SPADs have square-shaped active areas with rounded corners. To increase the fill factor, all pixels share one common deep N-well and hence the same cathode bias. Each SPAD has an active area of 24 μm × 24 μm, and there is a separation of ~14 μm between adjacent pixels, leading to a fill factors of ~45% for the array as a whole. Metal layers are placed on top of the inactive areas to prevent unwanted illumination penetration.

Figure 2(b) shows the electronic interface of a single SPAD. Breakdowns are read from the anodes, which are then connected to the TDC inputs through tristate buffers, and the shared cathode is constantly biased to a high positive voltage. Each SPAD can be quenched and loaded by forcing its anode to 3.3 V/0 V using a pair of NMOS and PMOS transistors.

Once an SPAD is loaded, it is left in a metastable state (floating anode), ready to be triggered. The SPADs are self-quenching. After a breakdown, the current flowing through the SPAD charges the capacitance at the anode with a time constant dependent on the dynamic resistance of the diode and the overall capacitance at the anode. As the anode capacitance is charged, the reverse voltage over the diode gradually decreases below the breakdown voltage, and consequently the current flow stops. With the necessary electronics placed outside the detector array to increase the fill factor, addition of the wiring capacitance roughly doubles the anode capacitance.
The crosstalk and delay differences of the electronics due to routing are static and were measured to be on the order of ~10 ps. They can be neglected in temporal measurement of the device crosstalk. This circuit-related crosstalk is not generally a problem and can be minimized via proper layout design.

The SPADs can be biased above breakdown during predefined time windows of up to 500 ns, whose width can be adjusted with a resolution of approximately 4 ns. The start and end of the time windows and their repetition (up to few hundreds of kilohertz) can be changed according to the application. After each time window and before the beginning of the next one, the measured breakdown time of all operative SPADs (with respect to a reference signal, e.g., the start of a laser pulse) are read out using a field-programmable gate array interface.

3. Noise-based crosstalk measurements and results

The measurements performed for the crosstalk assessment were “coincidence” measurements, in which detector SPADs are monitored after a breakdown in the emitter and the time differences of their breakdowns with respect to that of the emitter are recorded. In coincidence measurements, the temporal correlation of crosstalk events and the crosstalk probability can be evaluated directly, in contrast to “pseudo-crosstalk” measurements, in which the emitter is continuously reverse-biased above breakdown with a constant current for a specific integration time and the count rate of the detector is evaluated during this time [13–15]. “Pseudo-crosstalk” measurements have significantly shorter acquisition times; however, owing to their indirect nature, only a value proportional to the crosstalk probability is acquired, with no data on the timing relevance of crosstalk breakdown events.

In this study, the internal noise of the SPAD is used for triggering breakdown events in the emitter in coincidence crosstalk measurements. A noisy SPAD (with a DCR of 1.5 MHz)
was chosen as the emitter (E in Fig. 2) and left to be triggered by its own thermal or tunneling-generated dark counts, and the breakdowns of its two neighboring detectors (D1 and D2 in Fig. 2) were monitored. The SPADs were actively biased above their breakdown voltage for a time window of ~200 ns. In the case of a breakdown event in any of the SPADs within this time window, the triggered SPAD was self-quenched and remained inactive until it was activated again at the beginning of the next time window. This cycle was repeated with a rate of 100 kHz sufficiently to obtain nearly $10^9$ breakdown events in the emitter, leading to an acquisition time of approximately 10 h. Each time a detector was triggered after the emitter, its breakdown time difference with respect to the emitter breakdown was measured.

The DCRs of the detectors were measured (with the emitter inactivated) to be ~18 and ~13 kHz, respectively. Because the measurement time window (200 ns) is considerably shorter than the mean time interval between dark-noise avalanche occurrences (~56 and ~77 μs, respectively), it can be assumed that the own dark hits of the detectors are distributed evenly throughout the measurement window, meaning that the presence of the detectors’ own noise only introduces a constant offset to the number of hits measured per TDC bin. To obtain pure crosstalk hits, these evenly distributed dark-noise events were measured separately (with the emitter inactivated) and removed from the crosstalk measurements.

The measurement environment was kept consistently dark and at room temperature during the measurements, and the SPADs were operating with 3.3V excess bias. Figure 3(a) and 3(b) show the timing distribution of crosstalk events for D1 and D2 with respect to the emitter breakdown, and Fig. 3(c) shows the normalized timing distributions for D1 and D2 together. The overall crosstalk probabilities for D1 and D2 were measured to be 0.3% and 0.01%, respectively.

The crosstalk histograms for D1 and D2 have FWHMs of 700 and 400 ps and tails (90% to 10%) of ~1.5 and ~1.2 ns, respectively. The presence of a long tail in the crosstalk histograms and the wide FWHMs indicate electrical crosstalk in addition to optical. Both the direct and indirect optical paths from the emitter to any of the detectors have delays of < 1 ps. Thus, the optical-crosstalk distribution should represent the combined effects of the avalanche current build-up, spread, and quench in the emitter (i.e., the jitter of the emitter; ~100 ps), the relaxation time constant of the avalanche generated carriers (<1 ps), the added jitter of the detector (~100 ps), and the TDC uncertainty (~30 ps), which would lead to an FWHM of ~200 ps. The results suggest that the diffusion time of minority carries in the deep N-well due to electrical crosstalk contributes to the further widening of the distribution. On the other hand, the crosstalk-probability reduction from D1 to D2 is greater than the theoretically expected rate of direct optical attenuation $(1/r^2)e^{-αr}$, where r is the distance from the emitter, and $α$ is the absorption coefficient of silicon [9, 15]. This also suggests the presence of electrical crosstalk, which is attenuated with a higher rate proportional to $e^{-αr}$ [11]. The exponential shape of the crosstalk distribution in D1 indicates that electrical crosstalk has significant involvement, whereas for D2, a distinct combination of a Gaussian distribution (due to optical crosstalk) and an exponential distribution (due to electrical crosstalk) is evident.

Figure 3(d) shows the same normalized histogram for D2, together with the crosstalk events, in which D1 is triggered prior to D2. The main part of the crosstalk events (~90%) are found to be directly caused by E, as expected. The distribution of the crosstalk in D2—in which D1 is also involved—is delayed, has a wider distribution, and accounts for only 10% of the crosstalk events in D2. This distribution represents the convolution of the D1 crosstalk distribution with itself, and its exponential shape confirms the dominance of the electrical crosstalk in D1.
The indirect optical crosstalk due to the total internal reflection of secondary photons from the bottom of the chip (as measured in [14], for example) does not have a noticeable effect on the close neighbors of the emitter with the given SPAD dimensions of this design. Considering the silicon–air critical angle of 15° and substrate depth of nearly 700 µm, indirect crosstalk affects detectors with distances of approximately 375 µm from the emitter after the secondary photons have traveled approximately 1.5 mm in the substrate.

The noise-based crosstalk measurements were repeated for a few other detector/emitter pairs, confirming the aforementioned measurements, with the mean values of the crosstalk probability being 0.3% for D1 and 0.012% for D2.

4. Pulsed laser-based measurements and results

In another measurement, a short laser pulse from a semiconductor laser diode was directed through a single-mode fiber and focused using a microscope objective to illuminate the active area of the emitter (Fig. 4). The laser pulse had an FWHM of ~120 ps and a wavelength of ~870 nm [22] and was attenuated (using neutral-density filters) enough for the emitter to operate in the single-photon mode (detection rate of ~2%). Again, if detectors are triggered after the emitter breakdown, the detector breakdown delays with respect to the emitter breakdown were measured and recorded. The measurements were repeated with the same rate of 100 kHz to obtain emitter breakdown events (~10⁸), leading to an integration time of ~14 h. To ensure that any hits from the laser photons in the detectors (e.g., due to diffraction or objective non-ideality effects) were removed from the histogram—in addition to the detectors’ own dark-noise hits—a separate measurement was performed with the same setup but an inactive emitter and then used to remove the non-crosstalk hits.
Fig. 4. Schematic of the measurement system, with active illumination triggering the emitter.

Figure 5(a) shows histograms of hits versus time for E and D1 with respect to the laser-diode trigger pulse (histograms of hits). The emitter is triggered by a pulsed laser (870 nm) in the single-photon mode. To make the comparison between histograms in time domain easier, each histogram has been normalized with respect to its peak value. The emitter distribution FWHM reflects the laser pulse shape with the added jitter of the emitter SPAD and the TDC, leading to an FWHM of ~200 ps. The emitter histogram has a tail (90% to 10%) of ~0.9 ns, which is known to be caused by the diffusion of the generated holes due to photon absorption in the deep N-well beneath the p-n junction of the emitter to the multiplication region [23].

The probabilities of obtaining a correlated detection in D1 and D2 are measured to be 0.9% and 0.013%, respectively. This probability for D1 is considerably higher than the noise-based measurement result because not all the measured avalanches in D1 and D2 (in correlation with E triggering) are caused by crosstalk from other SPADs. Although the measurements are performed in the single-photon mode, this only means that with a high probability, the avalanche initiation in the E is caused by absorption of a single photon. On the other hand, there are multiple other photons from the laser pulse that might get absorbed in the deep N-well or substrate, which does not lead to E breakdown but might trigger detectors through the aforementioned processes. This correlated noise is not directly related to an emitter avalanche event (rather, it is related to the laser pulse arrival time at the emitter) but might be equally as important with regard to practical applications of pulsed-laser single-photon detection. However, the generated holes due to the absorption of laser photons have a low chance of diffusing to D2. As a result, no noticeable change is observed in D2.
Figure 5(b) shows the correlated noise distribution for D1 with respect to the E breakdown. The jitter of the laser pulse is not present in the distribution of Fig. 5(b) (in contrast to Fig. 5(a)). Considering the emitter histogram in Fig. 5(a) and the D1 correlated noise histogram of Fig. 5(b), both histogram tails are attributed to the diffusion of holes in the deep N-well, the former in the vertical direction and the latter in the lateral direction. The tail of the histogram in Fig. 5(b) (~1.7 ns) is twice as long as that of the emitter in Fig. 5(a) (~0.9 ns). The depth of the deep N-well being approximately 7 µm, suggests a maximum of $7 \times \sqrt{2}$ = ~10 µm diffusion distance for holes in the lateral direction, which is approximately equal to the minimum distance between the E and D1 active areas.

One advantage of triggering the emitter breakdowns with illumination is the possibility of investigating the correlated noise as a function of the light intensity and illumination wavelength. The pulsed-laser illumination intensity was measured at up to 20 × illumination power compared with the single-photon mode. The results show no significant difference in the correlated noise probability with the increase in illumination power. This is because of the proportional increase in both the trigger rate of the emitter due to the direct pulse illuminating it and the probability of the indirect triggering of detectors due to the laser pulse.

The pulsed laser-based measurements were repeated using a red pulsed laser (Hamamatsu C4725) with a wavelength of 630 nm. The resulting noise distribution is shown in Fig. 6, together with the previously shown distribution obtained via 870-nm laser measurement. Both noise distributions of Fig. 6 correspond to ~75 million emitter breakdowns.

![Fig. 6. Normalized correlated noise distribution for D1 for different incident illumination wavelengths.](image)

The probabilities of obtaining a correlated detection in D1 and D2 were measured to be 0.39% and 0.02%, respectively. The probabilities are smaller than those for the 870-nm laser, confirming the role of the absorption of laser photons in the deep N-well as a source of extra correlated noise. This can be explained by the shorter penetration depth of the red laser photons (~3 µm), which makes it more probable for the secondary laser photon absorption to affect the emitter rather than the detectors. There is no distinct difference in the timing distribution of the crosstalk events, which indicates that similar processes caused the noise events.

5. Discussion

Crosstalk has long been a major issue in hybrid SPAD-based detector arrays, in which separate SPAD and electronic dies are used to exploit custom technologies for improving the SPAD performance, including the detection efficiency and fill factor [24]. This yields a large extra parasitic due to bonding, which leads to a large avalanche current and consequently a
large crosstalk probability. Crosstalk is also gaining importance in CMOS SPAD arrays, as the density of pixels increases. In such arrays, to improve the fill factor and consequently the photon-detection efficiency (PDE) of the overall detector array, a common terminal is shared between many devices (usually the deep-N-well cathode). In some cases, to further increase the overall fill factor and PDE, interfacing electronics are placed outside the detector array (not in the exact vicinity of pixel). Thus, the parasitic capacitance at the anode (from where the breakdown events are read) is increased by the necessary long routes. Both of the above increase the probability of crosstalk.

As mentioned in the Introduction, there have been studies on the crosstalk characterization of SPADs, e.g., the important contribution of indirect optical paths [13, 14], the dependence of CMOS SPAD optical crosstalk on excess voltage, the distance between devices and die thickness [18], and the effect of the guard-ring width and p-n junction (multiplication region) depth on the crosstalk probability [25, 26]. However, few of these reports address the timing correlation of crosstalk events in CMOS arrays.

Nissinen et al. [12] conducted a noise-based coincidence crosstalk study on a 4 × 128 CMOS SPAD line array designed using 0.35-µm HV CMOS technology. On-chip TDCs with a resolution of 78 ps were used to evaluate the crosstalk-timing distribution but only during the first ~250 ps after the breakdown of the emitter. Wu et al. [20] presented timing-correlated crosstalk measurements on a CMOS SPAD array manufactured via a 0.18-µm HV CMOS process, in which a deep P-well and an N-isolation layer form the SPAD junction. The measurements showed a crosstalk distribution with an FWHM of ~250 ps corresponding to the combined jitter of the emitter and detector, indicating the non-existence of electrical crosstalk in the given structure with the center-to-center distance between the two measured devices exceeding 100 µm. In [11, 19], Vila et al. performed dark noise-generated coincidence crosstalk measurements on a very dense (67% fill factor) 5 × 1 CMOS SPAD array, counting crosstalk coincidences in minimum-2.5 ns time windows. The result emphasized the importance of the electrical crosstalk in structures with small spacing between SPADs.

Unlike other major noise sources in SPADs, such as dark and background noise (which have Poisson distributions), crosstalk is a correlated noise source. As a result, it can be easily mistaken with measured results in applications such as pulsed laser distance measurement. As indicated by the results of the noise-based coincidence measurements in a dense CMOS array, crosstalk events can occur with a probability of up to 1% with a timing distribution of a few nanoseconds. It is difficult to separate the observation of the optical and electrical crosstalk, but both contribute to the overall crosstalk.

In practice, when an SPAD is triggered by illumination, the correlated noise caused by the absorption of laser photons outside the multiplication region of the SPAD adds to the crosstalk. Owing to the low PDE of the SPAD, even under operation in the single-photon mode, many photons might not cause avalanches in the corresponding SPAD itself but might indirectly trigger avalanches in adjacent SPADs. This correlated noise is dependent on the wavelength of the illumination and the junction depth (as is the PDE), and in shallow junction SPADs in the infrared range, for example, can be an even more serious problem than crosstalk. In contrast to crosstalk, illumination-correlated noise exists even if the main SPAD is not triggered or is not biased above breakdown.

The most obvious solution for minimizing the crosstalk is increasing the distance between devices, which significantly reduces the fill factor [25, 26]. Another proposed solution for the crosstalk problem is the fabrication of deep isolating trenches (coated with metal [9, 27] or thick heavily doped deep diffusion regions [13, 14]) to avoid direct optical and electrical crosstalk (although this solution is still not prone to indirect optical paths). CMOS SPAD arrays do not have the freedom of introducing new fabrication process steps and thus remain vulnerable to crosstalk. Reducing the quench time of the triggered SPAD and hence the
number of carriers per avalanche is an effective method for reducing the crosstalk, but in 2D arrays, as the number of pixels increases, extra parasitic might set a limit on the reduction.

If the active area-to-pitch ratio of the SPAD array is increased (which is possible in this technology without premature breakdown occurring), the crosstalk problem can be even more extreme. Measures must be taken to compromise between the high fill factor and low crosstalk probability to ensure that the crosstalk does not harm the functionality of the receiver array.

**Funding**

Academy of Finland (Centre of Excellence in Laser Scanning Research, contract no. 307362).