# A Scaling Law for SPAD Pixel Miniaturization

**Kazuhiro Morimoto** [1,2] **and Edoardo Charbon** [1,*]

1   AQUA Laboratory, École Polytechnique Fédérale de Lausanne (EPFL), 2000 Neuchâtel, Switzerland
2   Canon Inc., Kanagawa 212-8602, Japan; morimoto.kazuhiro@mail.canon
*   Correspondence: edoardo.charbon@epfl.ch

**Abstract:** The growing demands on compact and high-definition single-photon avalanche diode (SPAD) arrays have motivated researchers to explore pixel miniaturization techniques to achieve sub-10 μm pixels. The scaling of the SPAD pixel size has an impact on key performance metrics, and it is, thereby, critical to conduct a systematic analysis of the underlying tradeoffs in miniaturized SPADs. On the basis of the general assumptions and constraints for layout geometry, we performed an analytical formulation of the scaling laws for the key metrics, such as the fill factor (FF), photon detection probability (PDP), dark count rate (DCR), correlated noise, and power consumption. Numerical calculations for various parameter sets indicated that some of the metrics, such as the DCR and power consumption, were improved by pixel miniaturization, whereas other metrics, such as the FF and PDP, were degraded. Comparison of the theoretically estimated scaling trends with previously published experimental results suggests that the scaling law analysis is in good agreement with practical SPAD devices. Our scaling law analysis could provide a useful tool to conduct a detailed performance comparison between various process, device, and layout configurations, which is essential for pushing the limit of SPAD pixel miniaturization toward sub-2 μm-pitch SPADs.

**Keywords:** single-photon avalanche diode; pixel miniaturization; scaling law

## 1. Introduction

Single-photon avalanche diodes (SPADs) have been widely recognized for having unique features, such as single-photon sensitivity and picosecond timing resolution. In recent decades, SPAD arrays fabricated with the silicon-based complementary-metal-oxide-semiconductor (CMOS) process have been extensively studied for a number of scientific and industrial applications. To explore the emerging applications of SPAD image sensors, researchers have developed large-scale SPAD pixel arrays in compact sensor formats. Continuous efforts in SPAD research and development have led to the exponential growth of the array size and dramatic shrinkage of the pixel dimension; the SPAD array size has reached a milestone of 1 megapixel [1], while a SPAD pixel with 2.2 μm-pitch was reported in test devices [2].

In addition, 3D-stacking approaches have enabled the physically isolation of pixel circuits from the SPAD array while ensuring electrical connection via pixel-level bonding, which provides a promising solution for pixel miniaturization below 10 μm [3–6]. Such an aggressive miniaturization and scaling of SPAD pixels could have a major influence on the key performance of SPADs, and it is, thereby, critical for designers to understand the fundamental tradeoffs in miniaturized SPADs. Theoretical studies on SPAD performance have been widely performed based on both analytical methods and simulations to describe process, voltage, and temperature dependence [7–9]. However, few attempts have been made to systematically analyze the impact of SPAD pixel scaling on major performance metrics, such as the fill factor (FF), photon detection probability (PDP), photon detection efficiency (PDE), dark count rate (DCR), correlated noise, and power consumption.

In this paper, we present an in-depth study of scaling laws in SPADs to clarify the underlying tradeoffs in SPAD design and to give perspectives for the future design of

multi-megapixel SPAD arrays. The formulation of scaling laws with regard to the pixel size is performed based on the assumption that the pixel circuit can be located outside of the SPAD array and does not impact the pixel layout. Based on the introduced equations, the scaling behavior of the SPAD performance is exemplified in the plots. The scaling law equations are then compared with previously published experimental results with various SPAD sizes. A good agreement between the theoretical fitting and experimental data validates the scaling law analysis.

The paper comprises four sections. Section 2 presents the theoretical formulation of scaling laws for various performance metrics in the SPAD pixels. Some examples of analyzing experimental data with the theoretical expressions is demonstrated in Section 3, and this is followed by discussion in Section 4.

## 2. Scaling Law Analysis

### 2.1. Analysis Criteria

To proceed with the theoretical analysis of the scaling laws, some assumptions must be made. First, the SPAD pixel array configuration is assumed to be a square grid, while it is not difficult to generalize the discussion to other configurations, e.g., honeycomb structure [10]. Second, circular-shaped SPADs are assumed to simplify the discussion on the curvature change with scaling. In some prior works, rounded-corner rectangle or square SPADs are also adopted to improve the fill factor [11–13]. However, these designs are not always suitable for scaling with the geometric similarity preserved, where the electric field concentration at the corners can induce premature edge breakdown and also change the breakdown voltage. Third, a 3D-stacked configuration with a SPAD-only array in a single plane is assumed.

In a non-3D-stacked FSI or BSI configuration, SPAD and pixel circuits coexist in the same plane. In a given pixel pitch, the SPAD and circuit have to share the limited area, and the circuit complexity can affect the size of the SPAD active area and its performance. The main focus of this analysis is to formulate the scaling laws of SPAD performance, and, hence, the SPAD array without circuit components is desired for more systematic and quantitative analysis. Fourth, the active-to-active distance is assumed to be fixed at a certain dimension, irrespective of the scaling parameter. This is justified by the following discussion.

For analysis of the scaling laws in the SPAD pixel, it is natural to assume that the doping profile along with the *z*-axis for each implantation layer is unchanged, and the breakdown voltage of the p-n junction in the SPAD remains in the same range. This implies that, unlike scaling in MOS transistors, where a lower supply voltage is adopted for the smaller devices, the power supply voltage for the SPAD does not scale as a function of its dimensions. Another premise in SPAD pixel design is that the guard-ring width is sufficiently large to avoid premature edge breakdown. Given the fact that the lateral diffusion length of doped ions cannot readily be controlled, the electrostatic potential distribution around the guard ring is not dependent on the active diameter. The optimum guard-ring width ensuring no edge breakdown in the operating condition is defined by the following equation [14]:

$$V_B^{gr}(W_{gr}) = V_B^{p-n} + V_{ex}^{max},$$ (1)

where $V_B^{gr}(W_{gr})$ is the breakdown voltage at the guard ring with the given guard-ring width $W_{gr}$, $V_B^{p-n}$ is the breakdown voltage at the vertical p-n junction, and $V_{ex}^{max}$ is the maximum excess bias used in the system. Based on the discussion above, all the terms in the above equation are not dependent on pixel size, and the optimum $W_{gr}$ can be defined regardless of scaling. These considerations impose a constraint in the pixel scaling that the guard-ring width has to be unscaled and fixed at a certain value over all the SPAD pixel dimensions to guarantee stable Geiger-mode operation without unwanted edge breakdown. The optimum $W_{gr}$ should be comparable to the depletion width of the main SPAD p-n

junction, and is typically 1 to 2 µm [15]. In addition, the optimum width of an isolation layer, typically formed with deep-well implantation, is determined by a process design rule for the minimum drawing width, and should not be scaled with the pixel dimensions. The pixel pitch $L_p$, which will be employed as a scaling parameter in the following discussion, can be expressed as:

$$L_p = D_a + L_{a-a} = D_a + 2 \cdot W_{gr} + W_{iso},　　　　　　　　　　(2)$$

where the well-sharing configuration is assumed, and $D_a$ is the active diameter, $L_{a-a}$ is the active-to-active distance, and $W_{iso}$ is the isolation width. In the following discussion, $W_{gr}$ and $W_{iso}$ are both assumed to be 1 µm unless otherwise noted, and $L_p$ is assumed to be solely dependent on $D_a$.

Figure 1 shows the conceptual views depicting the SPAD pixel scaling. Figure 1a is the example of a top-view layout for a 2 × 2 pixel array. As discussed above, the active-to-active distance $L_{a-a}$ is fixed when shrinking the pixel pitch $L_p$. As a result, the active diameter $D_a$ is reduced proportionally to $L_p$. This assumption can be applied to any type of existing SPAD device structures [16,17]. For example, Figure 1b shows the cross-sectional view of p+/NW SPAD. $D_a$ is defined as the diameter of the inner circle of the guard-ring p-well, whereas $L_{a-a}$ corresponds to a sum of the NW separation width and twice the width of the p-well guard ring. For PW/deep-NW SPAD or p-i-n SPAD, $D_a$ equals the diameter of the p-well, and $L_{a-a}$ is a sum of the NW separation width and twice the width of the virtual p-epi guard ring. This indicates that the scaling law analysis can be performed with only three key dimensional parameters, $L_p$, $D_a$, and $L_{a-a}$, without losing generality.

In summary, the main assumptions for the analysis of scaling laws are:

- a uniform square grid,
- a circular shape for the active area and inner/outer borders of the guard ring,
- a 3D-stacked configuration with full separation of the SPAD and pixel circuit into different wafers,
- an active-to-active distance unscaled with the SPAD pixel dimension, and
- the pixel pitch $L_p$ employed as a scaling parameter.
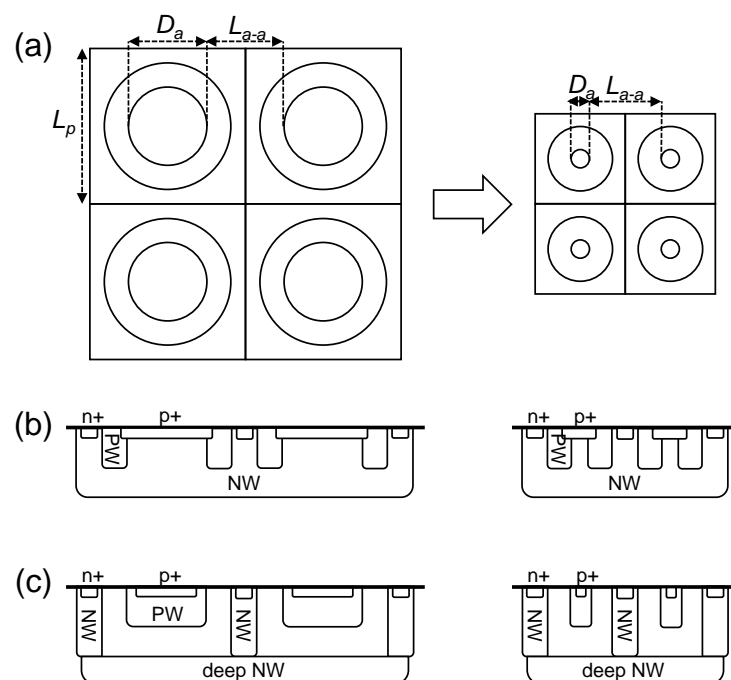


**Figure 1.** Conceptual views of SPAD pixel scaling; (**a**) top-view layout examples of pixel miniaturization, (**b**) cross-section example of p+/NW SPAD, and (**c**) cross-section example of PW/deep-NW SPAD or p-i-n SPAD.

*2.2. Formulation of Scaling Laws*

2.2.1. Fill Factor

The FF in the SPAD pixel, defined as the ratio between the drawn active area and the pixel area, is one of the fundamental parameters determining the single-photon sensitivity. FF is a purely geometric parameter, and is straightforward to be formulated as a function of the pixel pitch $L_p$:

$$FF = \frac{\pi(L_p - L_{a-a})^2}{4L_p^2}.$$

(3)

It is clear from the above equation that FF goes down to zero when $L_p = L_{a-a}$ and cannot be defined for $L_p < L_{a-a}$. For sufficiently large $L_p$, FF converges to $\pi/4 \times 100 = 78.5\%$.

Figure 2 shows the calculated FF as a function of the pixel pitch for several different active-to-active distances. FF curves show monotonic increases with the pixel pitch $L_p$. A relatively steep increase of FF is observed at smaller $L_p$, whereas saturating behavior of FF is shown at larger $L_p$. Slower saturation for larger $L_{a-a}$ indicates that, if the active-to-active distance is large, a larger pixel pitch is required to obtain a higher FF, e.g., above 50%. In the actual sensor design, the effective FF can be enhanced by employing on-chip microlenses [18,19], although designers should bear in mind that microlenses are less effective for smaller f-numbers of the main objective lens.
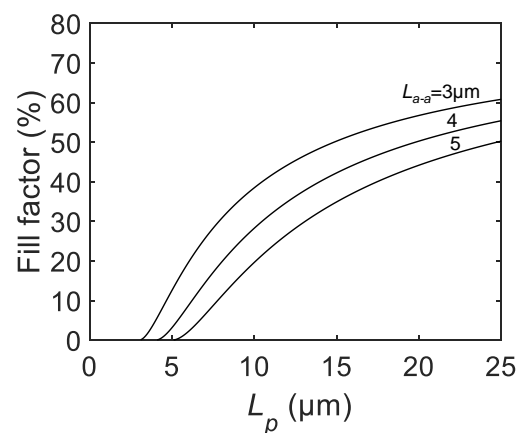


**Figure 2.** The calculated FF as a function of the SPAD pixel pitch $L_p$ for active-to-active distances $L_{a-a}$ = 3, 4, 5 μm.

2.2.2. PDP and PDE

The PDP in SPAD pixels is defined by the following equation [20]:

$$PDP = QE \times P_{ava},$$

(4)

where $QE$ is the quantum efficiency and $P_{ava}$ is the avalanche triggering probability. In ideal SPAD devices, PDP represents the single-photon sensitivity normalized by the active area, and it does not scale with the active diameter and the pixel pitch. In practice, a discrepancy between the "drawn" active area and "effective" active area leads to considerable dependencies of PDP from the scaling parameter $L_p$ [21].

The discrepancy between the designed and actual active size stems from two possible reasons: nonideality in the process fabrication and nonideality in the device design. One example of the process nonideality is the lateral diffusion of doped ions [22]. The lateral diffusion length is determined by the type of dopant ions, implantation energy, and thermal annealing conditions and is typically in the order of 0.1 to 1 μm for deep well implantation. This lateral diffusion induces the decrease of the doping concentration at the edge of the active area. The electric field at the edge of the active area can be locally reduced with

respect to the electric field at the center of the active area, thus, lowering the sensitivity at the border of the active area.

On the other hand, the device design nonideality is caused by a lateral electric field near the guard ring. Photocharges generated in the neutral region of the SPAD randomly move around due to thermal diffusion until they reach the nearby depletion region and are drifted to an electrode. If the photocharges reach the main p-n junction with a high electric field, they induce avalanche multiplication, thereby generating a photon detection signal. However, photocharges close to the border of the active area can reach the depletion region toward the guard ring before reaching the main junction. In such a case, the carriers do not cause avalanche multiplication, and no photon detection signal is observed. This so-called "border effect" [23,24] causes the photon detection loss at the edge of active area, which becomes more significant in the smaller pixels.

For both process- and device-originated nonidealities, PDP correction can be performed by introducing an inactive radius $r_{in}$, representing the effective width of the photon-insensitive region at the edge of the active region [25]. The corrected equation for the scaling law of PDP is given by:

$$PDP = PDP_{max} \times (\frac{L_p - L_{a-a} - 2r_{in}}{L_p - L_{a-a}})^2, \tag{5}$$

where $PDP_{max}$ is the virtual maximum PDP with a sufficiently large active size.

Figure 3 shows the calculated PDP as a function of the pixel pitch for different $r_{in}$. The curve with $r_{in} = 0$ μm corresponding to the ideal case with no border effect shows no dependency with $L_p$. For finite $r_{in}$, PDP starts from zero at $L_p = L_{a-a} + 2r_{in}$ and grows and saturates to $PDP_{max}$ with increasing $L_p$. Similar to the scaling law for the FF, a slower increase is observed for the larger $r_{in}$.
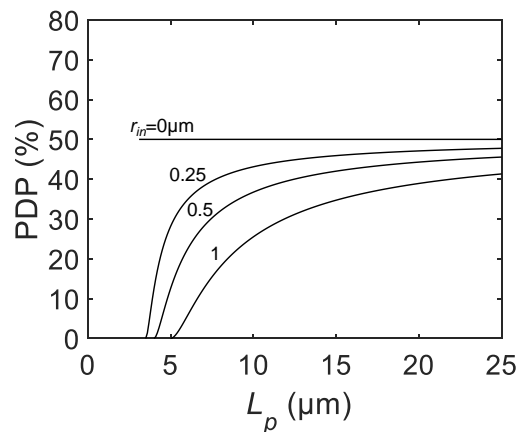


**Figure 3.** The calculated PDP as a function of the SPAD pixel pitch $L_p$ for $PDP_{max} = 50\%$, active-to-active distance $L_{a-a} = 3$ μm, and inactive radius $r_{in} = 0$, 0.25, 0.5, and 1 μm.

PDE is another indicator of single-photon sensitivity. Unlike PDP, where the sensitivity is normalized by the active area, PDE is defined as the single-photon sensitivity normalized to the pixel area. The following equation holds [20]:

$$PDE = PDP \times FF. \tag{6}$$

Based on the previous equations, PDE can be explicitly formulated as:

$$PDE = PDP_{max} \times \frac{\pi(L_p - L_{a-a} - 2r_{in})^2}{4L_p^2}. \tag{7}$$

Figure 4 is the calculated PDE as a function of $L_p$ for different $r_{in}$. Similar to FF and PDP, the curves start from zero at smaller $L_p$ and saturate at larger $L_p$. The maximum PDE is given by $PDP_{max} \times 78.5\% = 39.3\%$, assuming $PDP_{max} = 50\%$. Again, introducing on-chip microlenses will potentially increase the overall PDE.
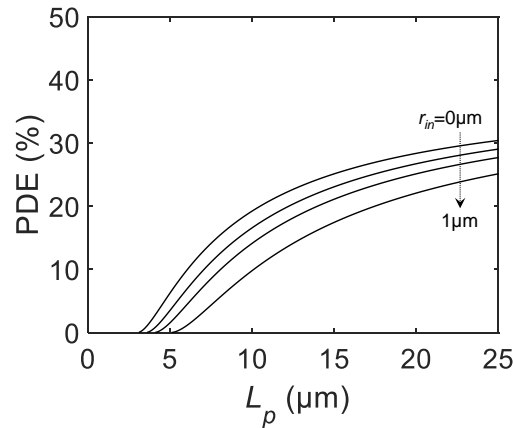


**Figure 4.** The calculated PDE as a function of the SPAD pixel pitch $L_p$ for $PDP_{max} = 50\%$, active-to-active distance $L_{a-a} = 3$ μm, and inactive radius $r_{in}$ = 0, 0.25, 0.5, and 1 μm.

### 2.2.3. DCR

DCR has several different causes, such as band-to-band tunneling, trap-assisted tunneling, trap-assisted thermal generation, and the diffusion current [26,27]. Experimentally, the source of the DCR can be classified based on an Arrhenius plot [28–30]. In silicon SPADs, the activation energies $E_a$ for band-to-band tunneling, trap-assisted tunneling, trap-assisted thermal generation, and the diffusion current are known to be approximately 0, 0–0.55, 0.55, and 1.1 eV, respectively. In practice, the measured $E_a$ can have intermediate values, e.g., 0.8 eV, indicating a mixture of multiple DCR components.

Based on the assumption that premature edge breakdown is suppressed, the tunneling components at the edge of the active region can be neglected. Contributions of the thermal generation and diffusion current are also negligible in the depletion region to the guard ring due to an insufficient electric field for avalanche triggering by the generated carriers. Therefore, the contribution from the main p-n junction of the SPAD dominates over that from the edge of the active region. Interestingly, all the aforementioned DCR components are proportional to the "effective" active area.

The tunneling current, regardless of being band-to-band or trap-assisted, is proportional to the total volume of the region with a highly concentrated electric field, which is clearly proportional to the active area. Thermal generation and diffusion carriers are detected only when those carriers are generated in the vicinity of the active region. Assuming that thermal generation and the diffusion current are spatially uniform around the active region, those components are also naturally assumed to be proportional to the active area. The scaling law for DCR can be formulated as follows:

$$DCR = R_0 \times \frac{\pi(L_p - L_{a-a} - 2r_{in})^2}{4}, \tag{8}$$

where $R_0$ is the DCR per unit of active area.

Figure 5 is the calculated DCR as a function of $L_p$ for different DCRs per unit area $R_0$. Starting from 0 cps at $L_p = L_{a-a} + 2r_{in}$, the DCR shows a parabolic increase with $L_p$. the DCR is highly dependent on $R_0$, which is a function of the excess bias, temperature, and process quality, such as the trap and impurity densities. Opposite to the FF, PDP, and PDE, a smaller pixel pitch is desirable to improve DCR performance. The designer should consider the best tradeoff between PDE and DCR to find the optimum $L_p$ to, thus, provide a reasonable S/N ratio.
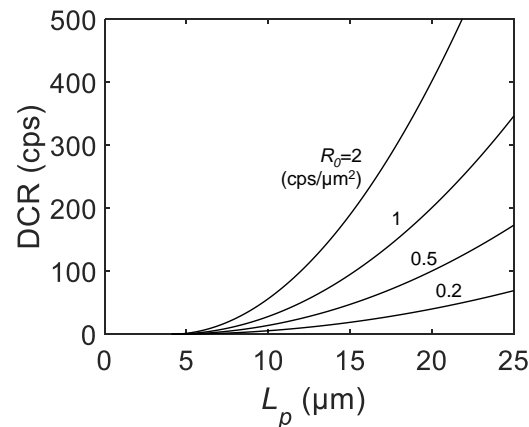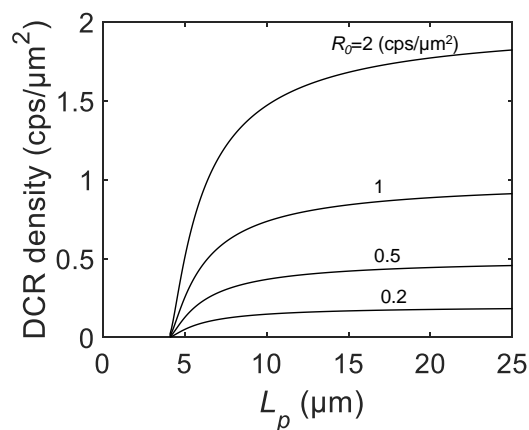
**Figure 5.** The calculated DCR as a function of the SPAD pixel pitch $L_p$ for $R_0 = 0.2$, 0.5, 1, and 2 cps/$\mu$m$^2$, active-to-active distance $L_{a-a} = 3$ $\mu$m and inactive radius $r_{in} = 0.5$ $\mu$m.

The DCR density $R$ is defined by the DCR normalized by the drawn active area and is often used for comparison of the SPAD process quality between devices fabricated in different processes [31]. As with PDP, nonideality, such as for the border effect, leads to the dependence of $R$ on $L_p$ as follows:

$$R = R_0 \times \frac{\pi(L_p - L_{a-a} - 2r_{in})^2}{4(L_p - L_{a-a})^2}. \tag{9}$$

At larger $L_p$, the DCR density saturates to $R_0$. Figure 6 shows the $L_p$ dependence of the DCR density for various $R_0$. As can be seen from the similarity to the equation for PDP, the DCR density starts from zero at $L_p = L_{a-a} + 2r_{in}$ and rapidly increases and saturates for larger $L_p$. This implies that, in the actual measurement, the DCR density can be underestimated at the smaller pixel pitch due to the existence of the photon-insensitive region at the edge of the active region.



**Figure 6.** The calculated DCR density as a function of the SPAD pixel pitch $L_p$ for $R_0 = 0.2$, 0.5, 1, and 2 cps/$\mu$m$^2$, active-to-active distance $L_{a-a} = 3$ $\mu$m, and inactive radius $r_{in} = 0.5$ $\mu$m.

Note that the above discussion is based on the assumption that the guard-ring width is optimized to avoid edge breakdown for the entire range of the pixel pitch. In the actual device design, sometimes an abrupt increase of DCR and DCR density is observed at a smaller pixel pitch even with fixed active-to-active distance. To the best of our knowledge, no systematic analysis has been conducted for this phenomenon. One possible reason is the enhanced curvature at the edge of the active region inducing a high electric field near the guard ring. Analogously to antennas, the electric field tends to increase in regions of

high curvature, which may induce premature edge breakdown when scaling down the pixel. Another possible explanation is the nonideality in the photoresist formation process.

In most SPAD devices, the diffusion regions for the p-n junction, guard ring, or isolation are formed by well doping where high energy doping is employed. In such a process, a thicker photoresist is desired to avoid penetration of the accelerated ions through the resist. The opening size of the photoresist for such a thick resist (typically 3 to 10 µm) requires careful calibration to match the actual shape and size to the designed layout. The layout for well doping is usually supported only for 0 or 90 degree lines, whereas a SPAD layout often involves a circular or ring shape with arbitrary angles. This could cause the deviation of the actual resist opening size from the design especially in the smaller pixel dimension, leading to unwanted edge breakdown.

### 2.2.4. Afterpulsing Probability

Correlated noise, such as afterpulsing and crosstalk, is critical for certain applications where the temporal and spatial correlations of photon detection signals play key roles [32–34]. Afterpulsing is caused by an avalanche-generated carrier captured at a deep trap state near the multiplication region, which is released by thermal activation or tunneling after a nanosecond to microsecond trapping time, thus, inducing another avalanche multiplication event. This mechanism implies that the afterpulsing probability $P_a$ is dependent on the trap density $D_{trap}$ and the total number of avalanche-generated carriers $N_{ava}$. A higher trap density and more avalanche carriers result in a higher $P_a$. If $P_a$ is not overly large, e.g., smaller than 10%, a linear relation between $P_a$ and $D_{trap} \times N_{ava}$ can be assumed to a first-order approximation [35].

Assuming the spatially uniform distribution of the deep trap states, $D_{trap}$ is independent of the scaling parameter. $N_{ava}$, on the other hand, can be dependent on the scaling parameter. $N_{ava}$ is calculated based on the following:

$$eN_{ava} = C_{par} \times V_{ex} = (C_{p-n} + C_0) \times V_{ex}, \tag{10}$$

where $e$ is the elementary charge; $V_{ex}$ is the excess bias; $C_{par}$ is the total parasitic capacitance at the SPAD output node, either cathode or anode, which is connected to the quenching resistor; $C_{p-n}$ is the p-n junction capacitance at the active region; and $C_0$ is the sum of the other parasitic capacitance contributions from connected metal wires, diffusion regions, gates, etc. $C_{p-n}$ is proportional to the active area, whereas $C_0$ does not scale with the pixel size or the active size. In summary, the scaling law of $P_a$ is given by:

$$P_a = A \times \left[ \frac{\pi\epsilon(L_p - L_{a-a})^2}{4W_{eff}} + C_0 \right], \tag{11}$$

where $A$ is the temperature-, bias-, and process-dependent coefficient; $\epsilon$ is the permittivity; and $W_{eff}$ is the effective depletion region width determined by the p-n junction doping profile.

Figure 7 shows the $L_p$ dependence of the afterpulsing probability for various $W_{eff}$ and $C_0$ (dashed lines for $C_0 = 5$ fF, and solid lines for $C_0 = 30$ fF). For all parameter combinations, the parabolic increase of $P_a$ is shown with the offset corresponding to $A \times C_0$. A larger $W_{eff}$ shows a weaker dependence of $P_a$ on $L_p$, indicating less contribution of the p-n junction capacitance to the total parasitic capacitance $C_{par}$. In any case, scaling down of the pixel has a positive impact on the afterpulsing probability due to the reduced parasitic capacitance.

Note that the dead time is assumed to be constant for all $L_p$ in this analysis. In a real device design, fixed quenching resistance results in the $L_p$ dependence of the dead time. This secondary effect makes the $P_a$ less sensitive to $L_p$ compared to the case where constant dead time is assumed. If the dependence of $P_a$ on the dead time is strong enough to compensate for the trend as shown in Figure 7 then it will be possible to flatten or even reverse $P_a$ for larger $L_p$.
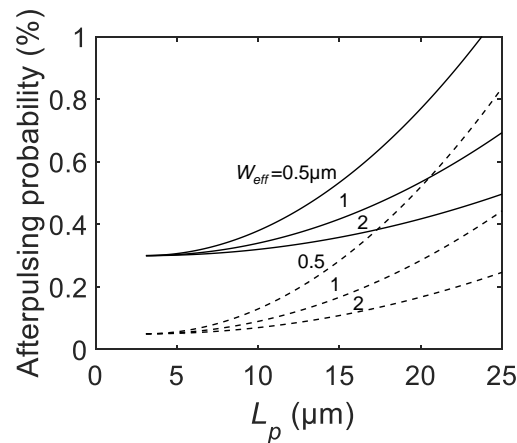
**Figure 7.** The calculated afterpulsing probability as a function of the SPAD pixel pitch $L_p$ for $C_0 = 5$ fF (dashed lines), and 30 fF (solid lines), active-to-active distance $L_{a-a} = 3$ μm, $A = 1 \times 10^{11}$ F$^{-1}$, and $W_{eff} = 0.5, 1, 2$ μm.

### 2.2.5. Crosstalk Probability

Crosstalk is another type of correlated noise in SPAD pixels. Unlike afterpulsing, where only a single pixel is involved, crosstalk involves two or more pixels. When avalanche multiplication is triggered in a pixel, thousands to millions of electrons and holes are generated. When those carriers are recombined with counterpart charges, either photons or phonons can be emitted to preserve the energy conservation law. Silicon is a material with an indirect bandgap, and hence the probability to emit photons is very low. For photon energy higher than the silicon bandgap, only several to tens of photons are emitted out of the one million avalanche-generated carriers [36,37]. However, those photons can move toward a neighboring pixel and be detected.

Similar to afterpulsing, the crosstalk probability $P_c$ is dependent on the number of avalanche-generated carriers $N_{ava}$. A larger number of carriers leads to a higher $P_c$. Again, to a first-order approximation, $P_c$ is considered to be proportional to $N_{ava}$. In addition, the distance between pixels is another important factor for scaling. Given that the emitted secondary photons decay exponentially with the travel length, a shorter pixel-to-pixel distance could result in higher crosstalk. The emitter-to-receiver distance dependence of crosstalk can be approximated by [38]:

$$P_c = B\frac{e^{-\alpha r}}{r^2},$$

(12)

where $B$ is a coefficient that will be explained later, $r$ is the distance from one SPAD of interest to the other, and $\alpha$ is the effective decay length of the emitted light. Regarding the crosstalk between two nearest-neighbor SPAD pixels, $r$ in the above equation corresponds to the pixel pitch $L_p$. Note that this equation implicitly assumes that the light emission occurs at the center of the active region for the emitting SPAD, and the average photon intensity reaching the active region of receiver is approximated by the photon intensity at the center of the active region of the receiving SPAD. In reality, the finite size of the active region for both the emitter and receiver may cause a slight deviation of the measured crosstalk from the above model. For simplicity, the following analysis will be based on the above model where the effect of the finite active size is neglected.

The coefficient $B$ is dependent on both the emitter and receiver characteristics. Considering the emitter, $B$ should depend on the total number of emitted photons, which is proportional to $N_{ava}$. On the other hand, $B$ should also be correlated with the sensitivity of the receiver. The probability of detecting an emitted photon is proportional to the PDP and the active area, which coincide with the PDE times $L_p^2$ by definition. Thus, the crosstalk probability between two nearest-neighbor SPAD pixels can be expressed as:

$$P_c = B' \times \left[ \frac{\pi \epsilon (L_p - L_{a-a})^2}{4 W_{eff}} + C_0 \right] \times \frac{e^{-\alpha L_p}}{L_p^2} \times \frac{\pi}{4} \times (L_p - L_{a-a} - 2r_{in})^2, \tag{13}$$

where $B'$ is an excess-bias dependent coefficient.

Figure 8 shows the $L_p$ dependence of the calculated crosstalk probability for various $\alpha$ and $C_0$. All curves show increasing trends for $L_p$ close to $L_{a-a} + 2r_{in}$. For larger $L_p$, either increasing or decreasing trends are observed, depending on the parameter set. The curve with $\alpha = 0.2, 0.1$ $\mu m^{-1}$, and $C_0 = 30$ fF shows reduction toward zero, whereas the curve with $\alpha = 0.05$ $\mu m^{-1}$ and $C_0 = 5$ fF shows a monotonical increase. Note that $\alpha = 0.2$ $\mu m^{-1}$, and $0.05$ $\mu m^{-1}$ correspond to the cases with effective light emission wavelengths of 700 and 850 nm, respectively. In contrast to afterpulsing, crosstalk probability does not necessarily show monotinic dependence on $L_p$; the impact of pixel miniaturization is highly dependent on the combination of model parameters.

To suppress crosstalk, several countermeasures can be considered. First, lowering $V_{ex}$ helps to suppress the crosstalk probability at the expense of the PDP and PDE. $V_{ex}$ affects both $N_{ava}$ in the emitter and the sensitivity of the receiver, and hence the crosstalk probability follows the square law with respect to $V_{ex}$. Second, the formation of opaque deep trench isolation (DTI) could suppress the crosstalk. Trench materials with a lower refractive index can reflect the emitted photons and eventually confine the photons in the emitter. This could lead to an order of magnitude improvement of the crosstalk probability.
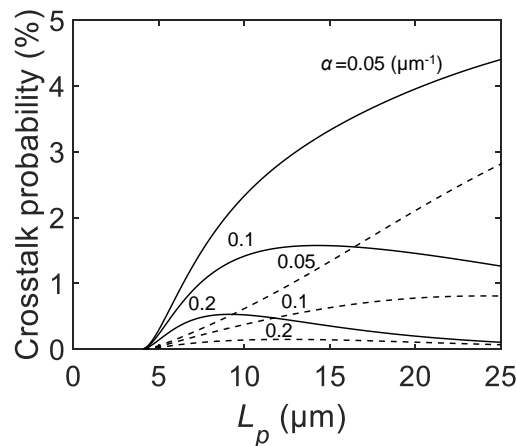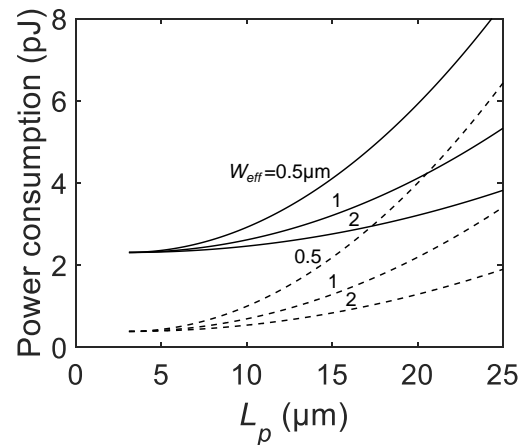


**Figure 8.** The calculated crosstalk probability as a function of the SPAD pixel pitch $L_p$ for $C_0 = 5$ fF (dashed lines) and 30 fF (solid lines), active-to-active distance $L_{a-a} = 3$ $\mu m$, $B' = 4 \times 10^{12}$ F$^{-1}$, $W_{eff} = 1$ $\mu m$, $r_{in} = 0.5$ $\mu m$, and $\alpha = 0.05, 0.1$, and $0.2$ $\mu m^{-1}$.

2.2.6. Power Consumption

The avalanche-originated power consumption in large-scale SPAD arrays is a key parameter as it grows proportionally to the number of pixels. The total power consumption in a SPAD array depends on the incident photon flux. For a systematic comparison, the following discussion focuses on the energy consumption per single avalanche event, $E_{ava}$, in a single SPAD pixel. The power consumption at the readout circuits is not taken into account here. $E_{ava}$ is a product of $eN_{ava}$ and $(V_{ex} + V_B)$, expressed as follows:

$$E_{ava} = C_{par} \times V_{ex} \times (V_{ex} + V_B) = D \times \left[ \frac{\pi \epsilon (L_p - L_{a-a})^2}{4 W_{eff}} + C_0 \right], \tag{14}$$

where $D = V_{ex} \times (V_{ex} + V_B)$ is the bias-dependent coefficient and $V_B$ is the breakdown voltage of the SPAD. Apart from the details of the coefficient, the equation has the same structure as that of the afterpulsing probability. Naturally, the calculated trend of the single-event power consumption $E_{ava}$ as a function of $L_p$ in Figure 9 shows similarity to Figure 7.

**Figure 9.** The calculated power consumption as a function of the SPAD pixel pitch $L_p$ for $C_0 = 5$ fF (dashed lines) and 30 fF (solid lines), active-to-active distance $L_{a-a} = 3$ μm, $V_B = 20$ V, $V_{ex} = 3.3$ V, and $W_{eff} = 0.5, 1,$ and $2$ μm.

### 2.2.7. Timing Jitter

The timing jitter in the SPAD is determined by multiple factors, such as the device configuration, doping profile, detection threshold, excess bias, and temperature, and it is not straightforward to formulate the scaling law for this. Qualitatively, a larger pixel pitch produces a higher timing jitter for several reasons: first, the spatial expansion of the avalanche multiplication process takes more time in the larger $L_p$ due to the finite lateral avalanche propagation velocity [39]. Second, a larger $L_p$ requires slower rising of the output voltage due to the larger parasitic capacitance, leading to enhanced statistical variability. Further systematic analysis should be conducted for deeper understanding of scaling the timing jitter.

### 2.2.8. Summary of Scaling Law Analysis

In the above sections, the scaling laws of the key SPAD characteristics with pixel dimensions were investigated. Miniaturization of the SPAD pixel improves the DCR, afterpulsing, power consumption, and timing jitter, whereas it has an adverse effect on the fill factor, PDP, and PDE. The equations for the scaling laws are summarized in Table 1. In particular, the degradation of the single-photon sensitivity is inevitable in the conventional SPAD pixel when its pitch becomes smaller than 10 μm. Further technological breakthroughs are required for SPAD pixel miniaturization toward multi-megapixel arrays.

**Table 1.** Summary of the scaling laws in the SPAD pixels with the pixel pitch $L_p$ as a scaling parameter. The coefficient is omitted in the equations.

| Characteristics | Equation |
|---|---|
| Fill factor (%) | $\frac{(L_p - L_{a-a})^2}{L_p^2}$ |
| PDP (%) | $(\frac{L_p - L_{a-a} - 2r_{in}}{L_p - L_{a-a}})^2$ |
| PDE (%) | $\frac{(L_p - L_{a-a} - 2r_{in})^2}{L_p^2}$ |
| DCR (cps) | $(L_p - L_{a-a} - 2r_{in})^2$ |
| DCR density (cps/μm²) | $\frac{(L_p - L_{a-a} - 2r_{in})^2}{(L_p - L_{a-a})^2}$ |
| Afterpulsing probability (%) | $\left[ \frac{\pi \epsilon (L_p - L_{a-a})^2}{4W_{eff}} + C_0 \right],$ |
| Crosstalk probability (%) | $\left[ \frac{\pi \epsilon (L_p - L_{a-a})^2}{4W_{eff}} + C_0 \right] \times \frac{e^{-\alpha L_p}}{L_p^2} \times \frac{\pi}{4} \times (L_p - L_{a-a} - 2r_{in})^2$ |
| Power consumption (pJ) | $\left[ \frac{\pi \epsilon (L_p - L_{a-a})^2}{4W_{eff}} + C_0 \right]$ |

## 3. Application to Experimental Results

*Extraction of Model Parameters*

To demonstrate the applicability of the scaling law analysis to practical situations, we performed a theoretical fitting with experimental results from the literature. Figure 10 shows experimental data from the literature [40] representing the pixel size dependence of the maximum PDP (shown as dots). Here, $L_{a-a}$ is assumed to be 8 μm. The experimental data was fitted using the scaling law equation for PDP (shown as a dashed line). The fitted curve shows a good agreement with the measurement when the fitting parameters are $PDP_{max}$ = 22.8% and $r_{in}$ = 1.06 μm.

The extracted fitting parameters indicate that the maximum PDP reaches 22.8% for larger pixels with this device configuration and bias condition, whereas the effective photon-insensitive region with the width of 1.06 μm reduces the maximum PDP for smaller pixels. The fitting result implies that the PDP will go down to zero at $L_p = L_{a-a} + 2 \times r_{in}$ = 10.12 μm, and thus the pixel pitch with this SPAD device configuration cannot go below 10 μm unless the process conditions and design rules are modified. Note that $r_{in}$ is determined by the spatial distributions of both the electrostatic potential and photon absorption rate. Given that the latter distribution is a function of the wavelength of incident photons, $r_{in}$ can potentially be dependent on the wavelength of interest.
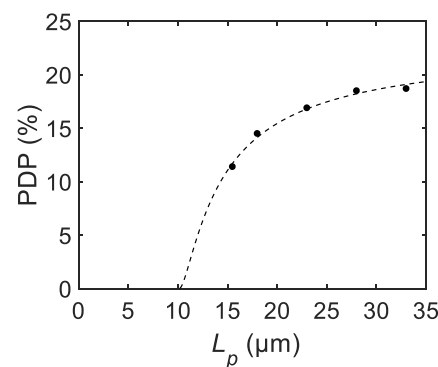


**Figure 10.** The measured PDP trend from the literature fitted by the theoretical equation [40]. The measured and fitted data are shown as dots and a dashed line, respectively.

In Figure 11, a similar analysis is performed for the measured DCR from the literature. Again, the fitting result shows good agreement with the measurement. The corresponding fitting parameters are extracted as $R_0$ = 8.50 cps/μm², and $r_{in}$ = 0.030 μm. An interesting implication is that the extracted $r_{in}$ for the DCR is different from that for the maximum PDP. This can be interpreted similarly to the previous remark on the wavelength dependence of $r_{in}$; the spatial distribution of the photon absorption rate for PDP can be different from that of the thermal generation rate for DCR, thereby, representing a different inactive radius $r_{in}$. The extracted $r_{in}$ for the DCR could provide useful information to estimate the major source of the DCR.
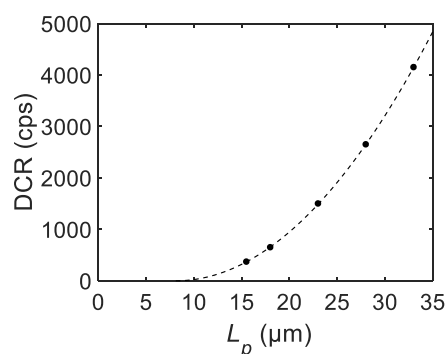
**Figure 11.** The measured DCR trend from the literature fitted by the theoretical equation [40]. The measured and fitted data are shown as dots and a dashed line, respectively.

## 4. Discussions

We investigated the theoretical expressions of the scaling laws for the major performance metrics in SPAD pixels. The analysis showed that SPAD pixel miniaturization improved the DCR, afterpulsing, power consumption, and timing jitter, whereas it had an adverse effect on the FF, PDP, and PDE. The scaling law equations for PDP and DCR were then applied to the experimental data in the literature, showing good agreement with the measured trends. The extracted fitting parameters were used to extrapolate the expected pixel size dependence of PDP and DCR, which implied that a pixel size smaller than 10 μm cannot be achieved without modification of the current process conditions and design rules.

Our scaling law analysis has three potential applications: the prediction of SPAD performance based on existing measurement data, extraction of model parameters to quantify the pixel-size-independent metrics, and systematic comparison of SPAD performance tradeoffs for different process, device, and layout configurations. The first approach can be useful for designers to understand the underlying tradeoffs and decide the optimal pixel size for applications of interest. The second approach can be critical to understanding the limiting factors of SPAD performance.

In-depth study of the extracted model parameters provides rich information of SPAD pixels, such as the inactive radius, parasitic capacitance, effective depletion width, and effective decay length of the avalanche-induced photons, which cannot be directly measured with existing measurement techniques. The third approach can be employed for clarifying the pros and cons of one SPAD device configuration to the other, which is essential for the correct choice of process conditions and device structure. Combining these approaches will provide a promising tool for further pushing the limit of SPAD pixel miniaturization toward sub-2 μm-pitch SPADs.

The extracted models are focused on pixel size dependence. Further generalization of the models to fully account for the voltage and temperature dependence of the metrics remains to be verified in future work.

## Abbreviations

The following abbreviations are used in this manuscript:

| | |
|---|---|
| CMOS | Complementary-metal-oxide-semiconductor |
| DCR | Dark count rate |
| DTI | Deep trench isolation |
| FF | Fill factor |
| PDE | Photon detection efficiency |
| PDP | Photon detection probability |
| QE | Quantum efficiency |
| SPAD | Single-photon avalanche diode |

## References

1. Morimoto, K.; Ardelean, A.; Wu, M.-L.; Ulku, A.C.; Antolovic, I.M.; Bruschini, C.; Charbon, E. Megapixel time-gated SPAD image sensor for 2D and 3D imaging applications. *Optica* **2020**, *7*, 346–354. [CrossRef] [PubMed]
2. Morimoto, K.; Charbon, E. High fill-factor miniaturized SPAD arrays with a guard-ring-sharing technique. *Opt. Express* **2020**, *28*, 13068–13080. [CrossRef]
3. Pavia, J.M.; Scandini, M.; Lindner, S.; Wolf, M.; Charbon, E. A 1 × 400 Backside-Illuminated SPAD Sensor with 49.7 ps Resolution, 30 pJ/Sample TDCs Fabricated in 3D CMOS Technology for Near-Infrared Optical Tomography. *IEEE J. Solid-State Circuits* **2015**, *50*, 2406–2418.
4. Al Abbas, T.; Dutton, N.A.W.; Almer, O.; Pellegrini, S.; Henrion, Y.; Henderson, R.K. Backside illuminated SPAD image sensor with 7.83 μm pitch in 3D-stacked CMOS technology. In Proceedings of the 2016 IEEE International Electron Devices Meeting (IEDM), San Francisco, CA, USA, 3–7 December 2016. [CrossRef]
5. Ximenes, A.R.; Padmanabhan, P.; Lee, M.-J.; Yamashita, Y.; Yaung, D.-N.; Charbon, E. A Modular, Direct Time-of-Flight Depth Sensor in 45/65-nm 3-D-Stacked CMOS Technology. *IEEE J. Solid-State Circuits* **2019**, *54*, 3203–3214.
6. Henderson, R.K.; Johnston, N.; Hutchings, S.W.; Gyongy, I.; al Abbas, T.; Dutton, N.; Tyler, M.; Chan, S.; Leach, J. A 256 × 256 40 nm/90 nm CMOS 3D-Stacked 120 dB Dynamic-Range Reconfigurable Time-Resolved SPAD Imager. In Proceedings of the 2019 IEEE International Solid-State Circuits Conference (ISSCC), San Francisco, CA, USA, 17–21 February 2019. [CrossRef]
7. Kindt, W.J.; van Zeijl, H.W. Modelling and Fabrication of Geiger mode Avalanche Photodiodes. *IEEE Trans. Nucl. Sci.* **1998**, *45*, 715–719. [CrossRef]
8. Gulinatti, A.; Rech, I.; Assanelli, M.; Ghioni, M.; Cova, S. A physically based model for evaluating the photon detection efficiency and the temporal response of SPAD detectors. *J. Mod. Opt.* **2011**, *58*, 210–224. [CrossRef]
9. Pancheri, L.; Stoppa, D.; Betta, G.-F.D. Characterization and Modeling of Breakdown Probability in Sub-Micrometer CMOS SPADs. *IEEE J. Sel. Top. Quantum Electron.* **2014**, *20*, 328–335.
10. You, Z.; Parmesan, L.; Pellegrini, S.; Henderson, R.K. 3 μm Pitch, 1 μm Active Diameter SPAD Arrays in 130nm CMOS Imaging Technology. In Proceedings of the 2017 International Image Sensor Workshop, Hiroshima, Japan, 30 May–2 June 2017. [CrossRef]
11. Maruyama, Y.; Blacksberg, J.; Charbon, E. A 1024 × 8, 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.
12. Niclass, C.; Matsubara, H.; Soga, M.; Ohta, M.; Ogawa, M.; Yamashita, T. A NIR-Sensitivity-Enhanced Single-Photon Avalanche Diode in 0.18 μm CMOS. In Proceedings of the 2015 International Image Sensor Workshop, Vaals, The Netherlands, 8–11 June 2015.
13. Gyongy, I.; Calder, N.; Davies, A.; Dutton, N.A.W.; Dalgarno, P.; Duncan, R.; Rickman, C.; Henderson, R.K. 256 × 256, 100 kfps, 61% Fill-factor Time-resolved SPAD Image Sensor for Microscopy Application. In Proceedings of the 2016 IEEE International Electron Devices Meeting (IEDM), San Francisco, CA, USA, 3–7 December 2016.
14. Gulinatti, A.; Ceccarelli, F.; Rech, I.; Ghioni, M. Silicon technologies for arrays of Single Photon Avalanche Diodes. *Proc. SPIE* **2016**, *9858*, 98580A.
15. Veerappan, C. Single-Photon Avalanche Diodes for Cancer Diagnosis. Ph.D. Thesis, TU Delft, Delft, The Netherlands, 2016. [CrossRef]
16. Richardson, J.A.; Webster, E.A.G.; Grant, L.A.; Henderson, R.K. Scaleable Single-Photon Avalanche Diode Structures in Nanometer CMOS Technology. *IEEE Trans. Electron Devices* **2011**, *58*, 2028–2035. [CrossRef]
17. Veerappan, C.; Charbon, E. A Low Dark Count p-i-n Diode Based SPAD in CMOS Technology. *IEEE Trans. Electron Devices* **2015**, *63*, 65–71. [CrossRef]
18. Pavia, J.M.; Wolf, M.; Charbon, E. Measurement and modeling of microlenses fabricated on single-photon avalanche diode arrays for fill factor recovery. *Opt. Express* **2014**, *22*, 4202–4213. [CrossRef] [PubMed]
19. Gyongy, I.; Davies, A.; Gallinet, B.; Dutton, N.A.W.; Duncan, R.; Rickman, C.; Henderson, R.K.; Dalgarno, P.A. Cylindrical microlensing for enhanced collection efficiency of small pixel SPAD arrays in single-molecule localisation microscopy. *Opt. Express* **2018**, *26*, 2280–2291. [CrossRef]
20. Charbon, E. Single-photon imaging in complementary metal oxide semiconductor processes. *Philos. Trans. R. Soc. A Math. Phys. Eng. Sci.* **2014**, *372*, 20130100. [CrossRef] [PubMed]

21. Takai, I.; Matsubara, H.; Soga, M.; Ohta, M.; Ogawa, M.; Yamashita, T. Single-Photon Avalanche Diode with Enhanced NIR-Sensitivity for Automotive LIDAR Systems. *Sensors* **2016**, *16*, 459. [CrossRef]
22. Tielert, R. Two-Dimensional Numerical Simulation of Impurity Redistribution in VLSI Processes. *IEEE Trans. Electron Devices* **1980**, *27*, 1479–1483. [CrossRef]
23. Acerbi, F.; Paternoster, G.; Gola, A.; Zorzi, N.; Piemonte, C. Silicon photomultipliers and single-photon avalanche diodes with enhanced NIR detection efficiency at FBK. *Nucl. Instrum. Methods Phys. Res. Sect. A* **2017**, *912*, 309–314.
24. Gola, A. NUV-HD and NIR-HD SiPMs and Applications. In Proceedings of the International SPAD Sensor Workshop, Les Diablerets, Switzerland, 26–28 February 2018.
25. Fishburn, M.W. Fundamentals of CMOS Single-Photon Avalanche Diodes. Ph.D. Thesis, Technische Universiteit Delft, Delft, The Netherlands, 2012.
26. Teranishi, N. Dark current and white blemish in image sensors. In Proceedings of the 2013 International Symposium on VLSI Technology, Systems and Application (VLSI-TSA), Hsinchu, Taiwan, 22–24 April 2013.
27. Charbon, E. Status and Future Prospects of High Time Resolution Photon Counting Sensor Arrays. In Proceedings of the Topical Workshop on Electronics for Particle Physics, Lisbon, Portugal, 28 September–2 October 2015. [CrossRef]
28. Webster, E.A.G.; Richardson, J.A.; Grant, L.A.; Renshaw, D.; Henderson, R.K. A Single-Photon Avalanche Diode in 90-nm CMOS Imaging Technology with 44% Photon Detection Efficiency at 690 nm. *IEEE Electron Device Lett.* **2012**, *33*, 694–696. [CrossRef]
29. 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*, 13200–13209. [CrossRef]
30. Xu, H.; Pancheri, L.; Betta, G.-F.D.; Stoppa, D. Design and characterization of a p+/n-well SPAD array in 150 nm CMOS process. *Opt. Express* **2017**, *25*, 12765–12778. [CrossRef]
31. 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*, 299–305. [CrossRef]
32. Unternährer, M.; Bessire, B.; Gasparini, L.; Stoppa, D.; Stefanov, A. Coincidence detection of spatially correlated photon pairs with a monolithic time-resolving detector array. *Opt. Express* **2016**, *24*, 28829–28841.
33. Gasparini, L.; Bessire, B.; Unternährer, M.; Stefanov, A.; Boiko, D.; Perenzoni, M.; Stoppa, D. SUPERTWIN: Towards 100 kpixel CMOS quantum image sensors for quantum optics applications. *Proc. SPIE* **2017**, *10111*, 101112L.
34. Lubin, G.; Tenne, R.; Antolovic, I.M.; Charbon, E.; Bruschini, C.; Oron, D. Quantum correlation measurement with single photon avalanche diode arrays. *arXiv* **2019**, arXiv:1910.01376. [CrossRef]
35. Liu, M.; Hu, C.; Campbell, J.C.; Pan, Z.; Tashima, M.M. Reduce Afterpulsing of Single Photon Avalanche Diodes Using Passive Quenching with Active Reset. *IEEE J. Quantum Electron.* **2008**, *44*, 430–434. [CrossRef]
36. Mönch, W. On the Physics of Avalanche Breakdown in Semiconducors. *Phys. Status Solidi* **1969**, *36*, 9–48. [CrossRef]
37. Snyman, L.W.; Plessis, M.D.; Aharoni, H. Injection-Avalanche-Based n+pn Silicon Complementary Metal–Oxide–Semiconductor Light-Emitting Device (450–750 nm) with 2-Order-of-Magnitude Increase in Light Emission Intensity. *Jpn. J. Appl. Phys.* **2007**, *46*, 2474. [CrossRef]
38. Rech, I.; Ingargiola, A.; Spinelli, R.; Labanca, I.; Marangoni, S.; Ghioni, M.; Cova, S. Optical crosstalk in single photon avalanche diode arrays: A new complete model. *Opt. Express* **2008**, *16*, 8381–8394. [CrossRef]
39. Assanelli, M.; Ingargiola, A.; Rech, I.; Gulinatti, A.; Ghioni, M. Photon-Timing Jitter Dependence on Injection Position in Single-Photon Avalanche Diodes. *IEEE J. Quantum Electron.* **2011**, *47*, 151–159. [CrossRef]
40. Malass, I.; Uhring, W.; le Normand, J.-P.; Dumas, N.; Dadouche, F. Evaluation of size influence on performance figures of a single photon avalanche diode fabricated in a 180 nm standard CMOS technology. *Analog Integr. Circuits Signal Process.* **2016**, *89*, 69–76.