A Simulation Model for Digital Silicon Photomultipliers

A Simulation Model for Digital Silicon Photomultipliers

Salvatore Gnecchi*†, Neale A.W. Dutton*†, Luca Parmesan* †,
Bruce R. Rae*, Stuart J. McLeod*, Sara Pellegrini*, Lindsay A. Grant*, Robert K. Henderson†
*ST Microelectronics Imaging Division, Edinburgh, United Kingdom
†The University of Edinburgh, Edinburgh, United Kingdom

***Abstract*—We propose a simulator model to estimate the performance of digital Silicon Photomultipliers (dSiPM) based on Single Photon Avalanche Diodes (SPADs) in terms of detection rate of photons incident on the sensor. The work provides guidelines for efficient array structure depending on: the number of SPADs, fill factor, area of both SPADs and array. A comparison of the main techniques present in the literature to digitally combine multiple outputs into single channel is included with simulated results showing promising higher detection rates for XOR-based dSiPMs. Mathematical expressions are derived to estimate dSiPM parameters such as maximum detection rate and detector dead time as functions of the mentioned design parameters.**

***Index Terms*—dSiPM, Silicon Photomultipliers, Single Photon Avalanche Diodes, SPAD, Model.**

I. Introduction

SILICON PHOTOMULTIPIERS (SiPM) have some advantages over Photomultipier Tubes (PMT) due to their insensitivity to magnetic fields, compactness and low voltage operation. A SiPM consists of an array of individual Single Photon Avalanche Diodes (SPAD) whose outputs are combined together to form a large area detector. There are two ways to combine the outputs of SiPMs: analogue SiPM (aSiPM) and digital SiPM (dSiPM), as illustrated in Fig. 1. In aSiPMs each avalanche current of the diodes is summed on a common node where the total current is then proportional to the number of photons detected by the sensor [1],[2]. The limitations of aSiPMs are: both the parasitic capacitance and the external load, e.g. the input stage of an ADC, which slow down the output signal, the cumulative dark noise and the high dependency of the output current to temperature. Moreover, at high light levels, aSiPMs show limited photon number resolution and they require external converters [3].

SPADs are amenable to be integrated in standard CMOS technology allowing CMOS circuits to be embedded on the same chip. SPADs, with a buffer integrated inside the pixel, are intrinsically digital photodetectors [4]. Such buffers and on-chip logic in dSiPMs overcome the dependency on the current value and the limitation on the photon number resolution thanks to precise counting and timing circuits (Fig.1(b)). For such reasons, dSiPMs are valid options in

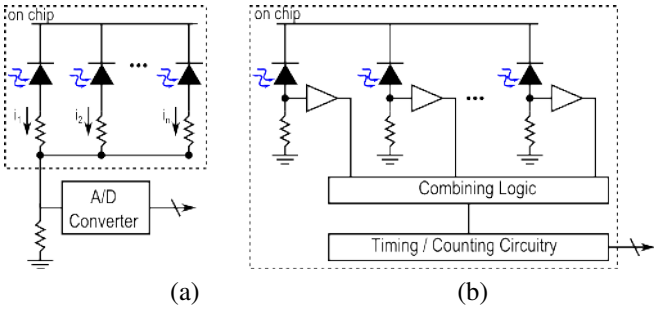


Figure 1. **Silicon Photomultipliers - SiPM** - SPADs outputs are combined in (a) analogue or (b) digitally techniques to realise higher detection rate sensors

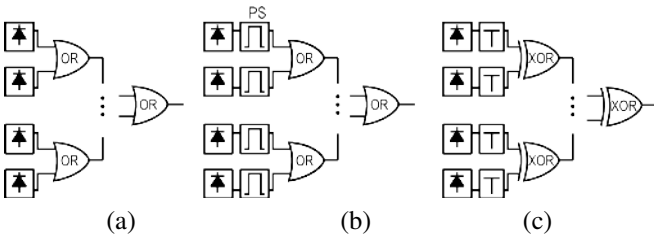


Figure 2. **dSiPM** - Digital ways to combine SPAD outputs: (a) OR tree, (b) Pulse Shortener + OR tree, (c) Toggle + XOR tree

a variety of applications such as Positron Emission Tomography (PET) [5], [6], Time Of Flight (TOF) ranging [7] and Fluorescence Lifetime Imaging Microscopy (FLIM) [8]. Several examples in the literature suggest three typical ways to combine the digital output of each SPAD pixel: OR tree [9], a Pulse Shortener (or Monostable) for each SPAD followed by an OR tree [8], [10] and a Toggle Flip-Flop per SPAD followed by an XOR tree [11] (Fig. 2).

In dSiPMs with embedded timing or counting circuitry, the output is affected by multiple sources of the pile-up effect, i.e. the reduction of registered events, due to the incapacity of certain components to process the quantity of incident photons. Arlt *et al.* have distinguished three main types of pile-up [12]. Detector pile-up is due to the SPAD dead time: the detector loses information of multiple photons incident on each single SPAD within its dead time. Secondly, routing pile-up is due to the combination of single outputs in

a dSiPM: the overall dead time of the detector will depend on the number of SPADs, on the SPAD dead time and on which combination logic is chosen. Finally, conversion pile-up describes the loss of information in the conversion step, mainly due to conversion dead time or limited conversion rate.

Photon pile-up leads to distortions of the output signal affecting the output photon statistics. For applications where the incident photon rate is very high ($\sim 10^9$ photons/s), such as close range proximity detection, pile-up represents a strong limitation. The typical solution to this problem is to adjust the intensity of the light source, where possible, thus reducing the number of incident photons at the price of losing photon dynamic range. A better approach is to match the detection rate with the conversion rate to minimise the pile-up.

Recently, converter pile-up has been addressed by high throughput Time to Digital Converters (TDCs) [11],[13]. The high conversion rate of TDCs leads to the main open question of how to efficiently match the speed of the conversion to the detection rate of the SiPM. In a dSiPM there are a number of parameters that need to be efficiently chosen and considered during the design: numbers of pixels, SPAD dead time, combining logic style, area and fill factor of the pixel and total area of the detector.

Existing models in the literature have considered the analogue aspects of SiPMs. Single SPAD models have been proposed in several works [14], [15],[16] focusing on the voltage and current characteristics of the SPAD output well describing parasitic effects and quenching techniques. For aSiPMs, works such as [17], [18] and [19] provide good examples of models which analyse the aspects of parasitics in a larger scale of a SPAD array. A model to analyse the response of an analogue SiPM under different light levels and number of fired pixels is presented in [20].

In this paper we examine dSiPM modelling: we provide a practical dSiPM simulation model with programmable parameters. Moreover, the model includes choices among the three main digital combination logic style. Simulations are presented to identify the highest detection rate of dSiPMs with a selected variety of the parameter choices. This information is useful to match the input rate of a signal converter thus minimising the overall pile-up effect.

Section II introduces in more details the digital SiPM whereas section III presents the MATLAB model for a single SPAD providing analysis for both active and passive recharge quenching circuit. Section IV contains the extension of the model to a full digital SiPM investigating the limitations to the maximum achievable signal rate offering comparisons using simulations results. The final section V includes conclusions in the general view of dSiPM detectors used in high photon rate applications and further outlook.

II. Digital Silicon Photomultiplier

The unit cell of a dSiPM is composed of the photodiode (SPAD) and the electronics (quenching circuit with readout buffer). Most commonly, each pixel also includes pull-up or pull-down transistors for enable/disable of a region of interest of the array to enable the rejection of high DCR pixels without significantly affecting the fill factor. To simplify the routing in dSiPMs with large number of SPADs, combining logic can also be included in the pixel. Fig. 3 illustrates the three main techniques used in dSiPM design. The use of a Pulse Shortnener (PS) is a common technique [8],[7] and provides immediate improvement compared to the simple OR tree combination in terms of signal rate: combining less wide pulses allows the final signal to reach higher frequencies. In a certain time window, the number of rising edges divided by the length of the time window gives an estimation of the average detection rate of the sensor. In this case, the frequency of the signal is no more limited by the SPAD dead time but it is determined by the length of the pulse shortener. As proposed by Dutton *et al*. in [11], SPAD pulses can be combined together without the need of a pulse shortener but implementing a toggle cell in the pixel to be then followed by an XOR tree replacing the OR tree. The elimination of the pulse shortener removes the need of routing its control voltage down to the pixel level and more important it allows overcoming the limitation due to the length of the output pulse. Fig. 3 also shows an important consideration that needs to be taken into account when comparing these last two approaches. A crucial difference between OR tree and XOR tree becomes essential for high incident photon rate. In fact, two or more very close in time photons that are successfully detected by two or more different SPADs are merged together by the OR tree making only the first photon survive the detection process. All the rest of the photons are piled-up and their information is lost. When the same scenario happens with an XOR tree though, all the photons are lost due to the limitation of the logic to create fast consequent transitions in the final signal. Very close in time photons are therefore all lost. The aim of this model is also to understand how to compare the benefits of the Toggle + XOR tree against the PS + OR tree and the disadvantage of the potential loss of the first photon.

A general simulation flowchart is shown in Fig. 4(a). First, initial parameters are chosen for the analysis, Table I. The simulation step time t_{step} is chosen to be 1ps in order to reach good accuracy in the time resolution of the SPAD output. The time window T is arbitrarily chosen to be $5\mu\text{s}$ to trade-off accuracy and simulation time. Example values of SPAD parameters are chosen ($PDE_0 = 0.6$, dead time $\tau_{\text{dead}} = 10\text{ns}$, excess bias $V_e = 1.2\text{V}$) and are easily modified to suit specific SPAD structures. The photon rate n is swept for all the simulations in a useful range to show the maximum of the output signal rate. The number of SPADs N_{SPADs} is swept following the same principle.

According to the light level described by the incident photon rate n, the total number of photons N in the time window T is calculated as $N = n \times T$. The photons are distributed in the time window according to a Poisson distribution, see Fig. 4(b). The vector containing the time distributed photons

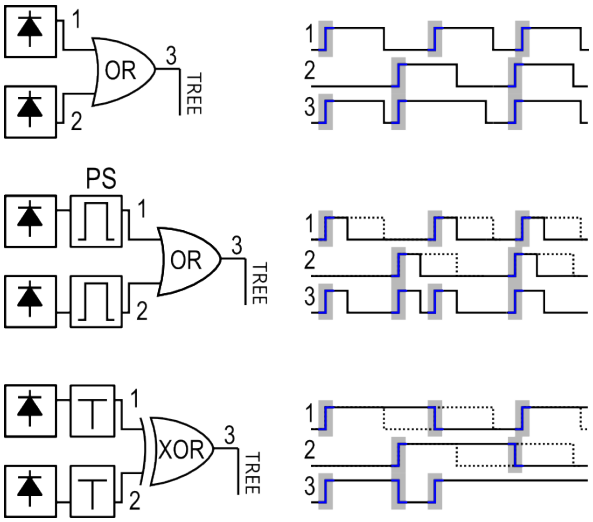
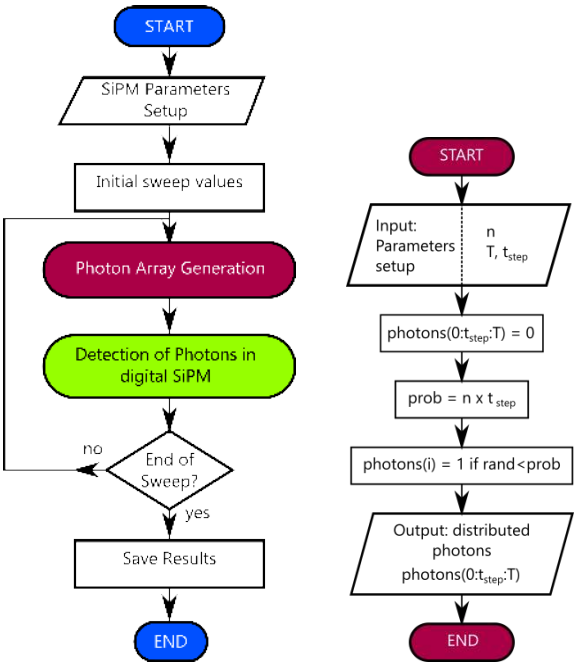


Figure 3. **Digital Silicon Photomultiplier and pile-up** - Three main techniques with examples of photon pile-up: OR-tree logic, pulse shaper + OR-tree logic (SPAD pulse = dashed lines, pulse shaper output = solid lines), Toggle + XOR-Tree Logic (original pulse = dashed lines, toggle output = solid line)



(a) (b)

Figure 4. **General Simulation Flow** - A typical simulation (a) sees the generation of photons (b), given a light level, and the detection of the photons. The routines are repeated changing light level and/or number of SPADs in the digital SiPM

Table I
PARAMETER SETUP

Parameter	Description	Chosen Value
T	Time window	5 µs
t_{step}	Simulation time step	1 ps
PDE_0	Photon Detection Efficiency	0.6
τ_{dead}	SPAD dead time	10 ns
V_e	Excess Bias	1.2 V
n	Incident Photon Rate	10^6 to 10^{11} Hz
N_{SPADs}	Number of SPADs	1 to 1000
τ_{PS}	Output Pulse Shortner Width	100 ps
$\tau_{\text{PW,min}}$	Minimum Pulse Width	100 ps

(0= no photon, 1= photon) and the simulation parameters are then passed to one of the algorithms for the photon detection of the digital SiPMs described in the following sections. For each sweep step, the output is scanned to count the number of transitions M (only rising edges in the OR case, both rising and falling in the XOR case) and calculate the average detection photon rate as $m = M/T$. The results are then analysed to derive modelling equations.

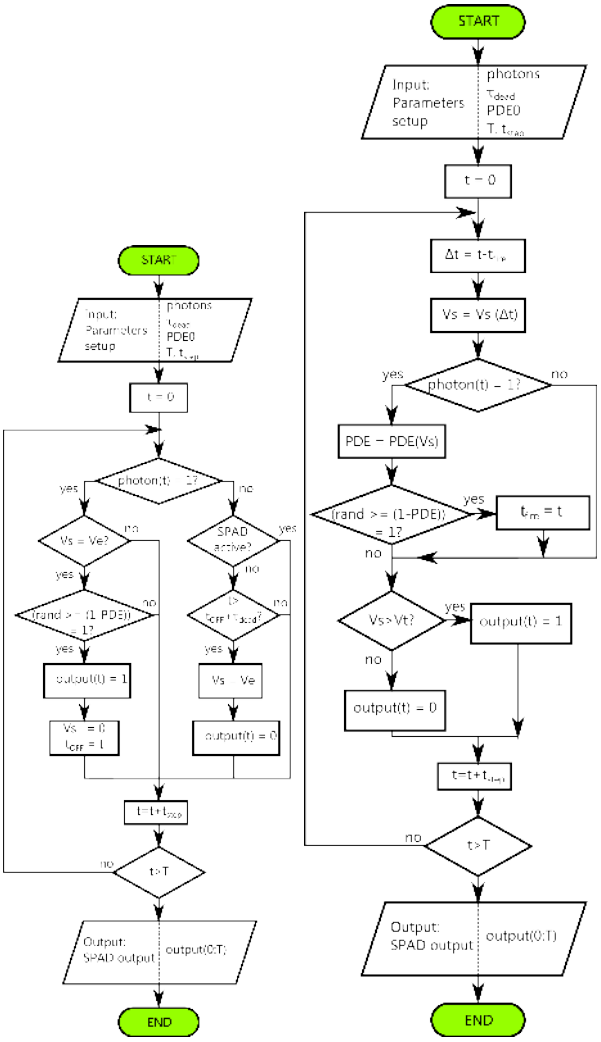
III. SINGLE SPAD MODEL

In this section two models for the single SPAD detector are presented and applied to analyse the signal rate out of the SPAD pixel. The first of the two models simulates an active recharged SPAD where the dead time τ_{dead}, i.e. the time window where the SPAD is inactive, is considered constant. The second one simulates a passive recharged SPAD for which the dead time increases at higher light levels because of the paralysis of the detector [21].

The algorithms provide, as output, a time sequence of digital 1s and 0s saved in a vector where every rising transition 0 to 1 represents a detection of a photon in the time window described by the time step of the vector. In this assumption, the falling edge 1 to 0 does not contain any time information.

A. Active Recharge SPAD

The algorithm to decide if and when a transition occurs is shown in Fig. 5(a). At each time step, the function checks if a photon has hit the SPAD. If yes, the triggering happens if the SPAD voltage V_S is correctly set to an excess bias V_e, assumed constant and equal to V_e, see Table I. Moreover, the PDE is modelled by the generation of a uniformly pseudo-random number between 0 and 1 and by comparing that to the assumed $PDE = PDE_0$. In case of photon conversion, the SPAD is finally triggered, the output vector is set to 1 and the dead time window begins: the time stamp of the detection is saved in a temporary variable t_{OFF} and for a period of τ_{dead} the SPAD is kept to $V_S = 0$, inhibiting the possibility of detecting photons. For all the dead time, the output vector is held to 1. The SPAD voltage returns to the excess bias V_e when the dead time window is completed. The output vector is reset to 0.



(a) (b)

Figure 5. **SPAD detection routine** - Algorithm for photon detection of a single SPAD in a certain time window - active recharging (a) and passive recharging with paralysis effect (b).

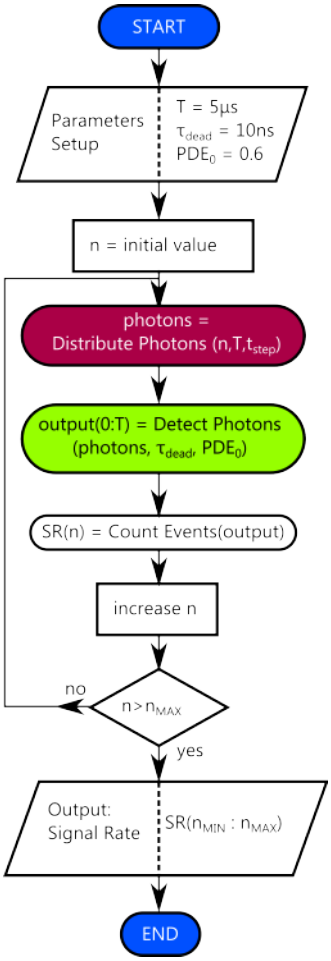


Figure 6. **Single SPAD detector Simulation**: The detection of the photons is called for different light levels (photon frequencies). The signal rate is calculated by counting the converted events in the chosen time window

The simulated algorithm is shown in Fig. 6. The signal rate is plotted against the photon rate in Fig. 7. The average signal rate $m = M/T$ follows the equation:

$$m = \frac{1}{\tau_{\text{dead}}} \left(1 - e^{-PDE \times n \times \tau_{\text{dead}}}\right) \quad (1)$$

For low values of n the signal rate is not affected by the limit of the dead time:

$$m \sim \frac{1}{\tau_{\text{dead}}} \left(1 - (1 - PDE \times n \times \tau_{\text{dead}})\right) = PDE \times n \quad (2)$$

For higher n the signal rate is limited and converges to a maximum of $1/\tau_{\text{dead}}$. These results match experimental values from previous work by Eisele *et al.* [21].

B. Passive Recharge SPAD

The next model considers the implementation of the paralysis effect on passive recharged SPADs. The algorithm

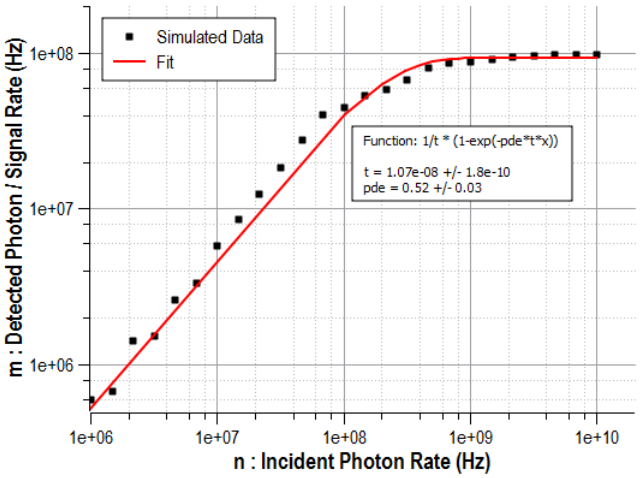


Figure 7. **Signal SPAD with Active Recharge** - The signal rate saturates at $1/\tau_{\text{dead}}$

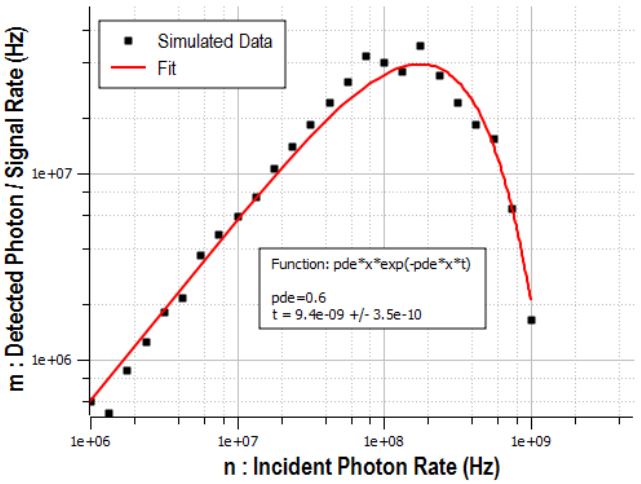


Figure 8. **Signal SPAD with Passive Recharge** - The signal rate reaches its maximum value m_{MAX} at $1/(\tau_{\text{dead}} \times PDE_0)$

has been modified as follows. The detection of the photon is now modelled according to Savuskan *et al.* [22] with a non-constant PDE depending on the SPAD bias voltage V_S:

$$PDE = PDE_0 \times (1 - e^{-V_S/V_0}) \tag{3}$$

The SPAD voltage V_S in passively recharged SPAD increases back to the set excess bias V_e with an RC characteristic:

$$V_S(\Delta t) = V_e \times (1 - e^{-\Delta t/RC}) \tag{4}$$

where Δt is defined as the time elapsed since the last detection. Even if the photon detection always resets the recharge process ($\Delta t = 0$), the output of the SPAD pixel goes back to zero only when the V_S exceeds a threshold value V_t (typically $V_t = V_e/2$). The effect of the reset of the recharge process has the consequence of increasing the effective dead time of the SPAD. The modified algorithm is illustrated in Fig. 5(b). The simulation has been kept equivalent to the previous case and results are shown in Fig. 8. The signal rate does not saturate as for active recharged SPADs. Instead, it reaches a maximum before rapidly decreasing. The behaviour is well described by a paralysed model [23] where the average signal rate m versus the hitting photon rate n is given by:

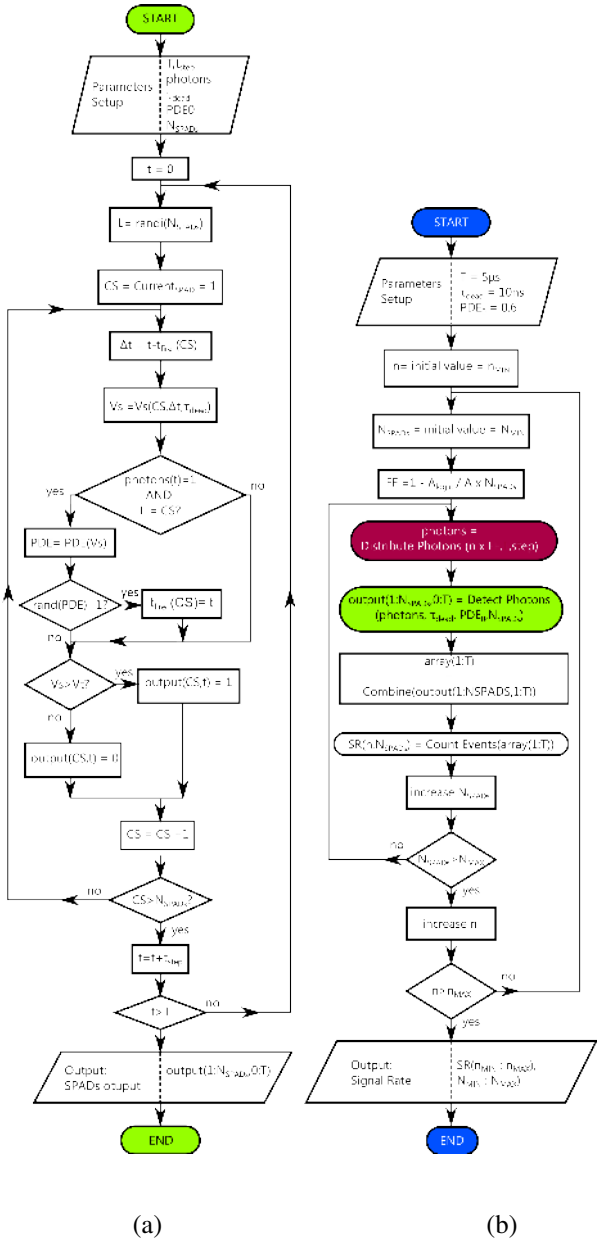
$$m = PDE_0 \times n \times e^{-PDE_0 \times n \times \tau_{\text{dead}}} \tag{5}$$

which has its maximum at:

$$n_{\text{MAX}} = \frac{1}{\tau_{\text{dead}} \times PDE_0} \Rightarrow m_{\text{MAX}} = \frac{1}{e \times \tau_{\text{dead}}} \tag{6}$$

IV. Digital SiPM Model

The model is now extended to describe the global performance of more SPADs digitally combined together into a single array forming a digital SiPM. In this way, the photo efficiency of the sensor is improved by the possibility of electrical events being triggered by more than a single SPAD.



(a) (b)

Figure 9. **Digital SiPM** - Generalised algorithm for an N_{SPADs} digital SiPM. Each photon is potentially detected by a SPAD chosen randomly in the array. Passive quench is considered.

The size of the array, i.e. the number of the aggregated SPADs in a constant area, is the first point of interest. The model helps understanding the limit at which having more SPADs does not improve the signal rate of the output of the array.

In case of a pulse shortener and OR tree, the detected photon rate has intrinsic limitation in the width of the pulse shortener output τ_{PS} and the gate delay of the subsequent electronics τ_{electr}. The alternative solution of toggle flip-flop and XOR tree removes the limitation of τ_{PS} leaving only the technology limitation of $\tau_{\text{PW,min}}$, i.e. the minimum signal pulse width. This section applies the developed simulator to understand the eventual trade-off between these two

solutions.

The improved algorithm, used to simulate the behaviour of an N_{SPADs} digital SiPM, is shown in Fig. 9. The assumption is that for each time step, only one photon is incident on one of the SPADs in the array. The time step needs therefore to be chosen small enough to see the saturation effect in the simulation results. This is here set to 1ps. The one SPAD is randomly chosen among the N_{SPADs}. A loop then scans each SPAD to look for the chosen one to eventually detect the photon or to recharge it in case of previous events. This is the generalisation of the single SPAD detection with passive quench of Fig. 5(b) (it can be used with $N_{\text{SPADs}} = 1$).

The main parameters in this model are: the number of SPADs N_{SPADs} and the rate of incident photons n. Both are swept in simulations and plots of the signal rate are presented. The algorithm is shown in Fig. 9(b). The following sections analyse different approaches of aggregating SPADs in the array by describing the impact on each parameter and the effects on the final signal rate of the array.

A. Fixed SPAD Area

In digital SiPM design, we study how many SPADs with the same active area A_1 and dead time τ_{dead} are needed to reach a maximum signal rate. When N_{SPADs} are chosen, the total light sensitive area A_N increases linearly with the number of diodes. At a constant light level n, a larger array will therefore collect more light thus improving the incident photon rate (7).

$$n_{\text{total}} = n \times N_{\text{SPADs}} \tag{7}$$

The algorithm has been run with the same setting as before and the results are presented in Fig. 10. The maximum signal rate suggested by the results of the model is:

$$m_{\text{MAX}} = \frac{1}{e \times \tau_{\text{MIN}}} = \tag{8}$$

$$\frac{1}{e \times (\tau_{\text{PS}} + \tau_{\text{electr}})} \quad \text{, for OR-tree} \tag{9}$$

$$\frac{1}{e \times \tau_{\text{PW,min}}} \quad \text{, for XOR-tree} \tag{10}$$

The model suggests that the N_{SPADs} dSiPM can be seen as a single SPAD with effective dead time $\tau_{\text{dead,eff}}$, effective incident photon rate n_{total} depending on N_{SPADs}. The effective incident photon rate depends on the number of SPADs through (7). The effective dead time for the whole array is now investigated. For a single SPAD the dead time is known and equal to τ_{dead}, hence:

$$\tau_{\text{dead,eff}}(N_{\text{SPADs}})\Big|_{N_{\text{SPADs}}=1} = \tau_{\text{dead}} \tag{11}$$

For a very large number of SPADs, the SPAD array behaves as its effective dead time is lowered to a minimum number given by:

$$\tau_{\text{dead,eff}}(N_{\text{SPADs}})\Big|_{N_{\text{SPADs}}\to N_{\text{Best}}} = \tau_{\text{min}} = (\tau_{\text{PS}} + \tau_{\text{electr}})$$

$$\text{, for OR-tree array} \tag{12}$$

$$= \tau_{\text{PW,min}}$$

$$\text{, for XOR-tree array} \tag{13}$$

The simplest way to describe such limits is the following function:

$$\tau_{\text{dead,eff}}\big(N_{\text{SPADs}}\big) = \tau_{\text{min}} \Big/ \left(1 - e^{-N_{\text{SPADs}} \times \tau_{\text{MIN}} / \tau_{\text{dead}}}\right) \tag{14}$$

Then (7) and (14) are used to modify (5). As result, the following equation can be considered valid for a SPAD array under the described assumptions:

$$m = PDE \times N_{\text{SPADs}} \times n \times e^{-N_{\text{SPADs}} \times n \times PDE \times \tau_{\text{dead,eff}}} \tag{15}$$

To compare the results of the simulator to the predicted modelling equation (15), the signal rate m has been plotted against the photon rate n for different values of N_{SPADs} (hence different values for the effective dead time through (14)). Results of the comparison are shown on the right side of Fig. 10. The graphs show that for a number of SPADs around $N_{\text{SPADs}} \sim 100$ (see (16) for calculation) the average signal rate is higher than the predicted paralysable model. The paralysis effect, equation (15), is not dominant for such number of SPADs: the SiPM has a hybrid behaviour between the non-paralysed detector (dead time not dependent on the photon rate) and paralysed detector (dead time increasing with high photon rate). After this region, the number of SPADs gets *too high*: the very close in time photons are not sufficiently detected due to the paralysis of the SiPM. The results of this simulation confirms that the collecting more light by using a larger SiPM helps in terms of number of SPADs but only up to a certain value $N_{\text{SPADs,best}}$ after which the SiPM, at high light level, either holds a high signal rate in a hybrid behaviour, or shows a lower detection rate due to pure paralysis.

In Fig. 11 the effective dead time predicted by (14) is compared to the point and value of maximum signal rate against the number of SPADs in the array. The data and the equation agree in the prediction of the number of SPADs needed to reach the maximum signal rate: $N_{\text{SPADs,best}} \simeq 100$. Assuming that for a low number of SPADs the effective dead time decreases linearly, then $\tau_{\text{dead,eff}} \sim \tau_{\text{dead}} \times N_{\text{SPADs}}$ predicts the optimum number of SPADs as:

$$N_{\text{SPADs,best}} = \frac{\tau_{\text{dead}}}{\tau_{\text{min}}} \tag{16}$$

which equals 100 in the simulated case. Fig. 11, together with the results of Fig. 10, shows that choosing a number of SPADs greater than $N_{\text{SPADs,best}}$ does not only give an improvement in the signal rate but also increase the paralysis effect. Under these assumptions, the most efficient number of SPADs is predicted by (16).

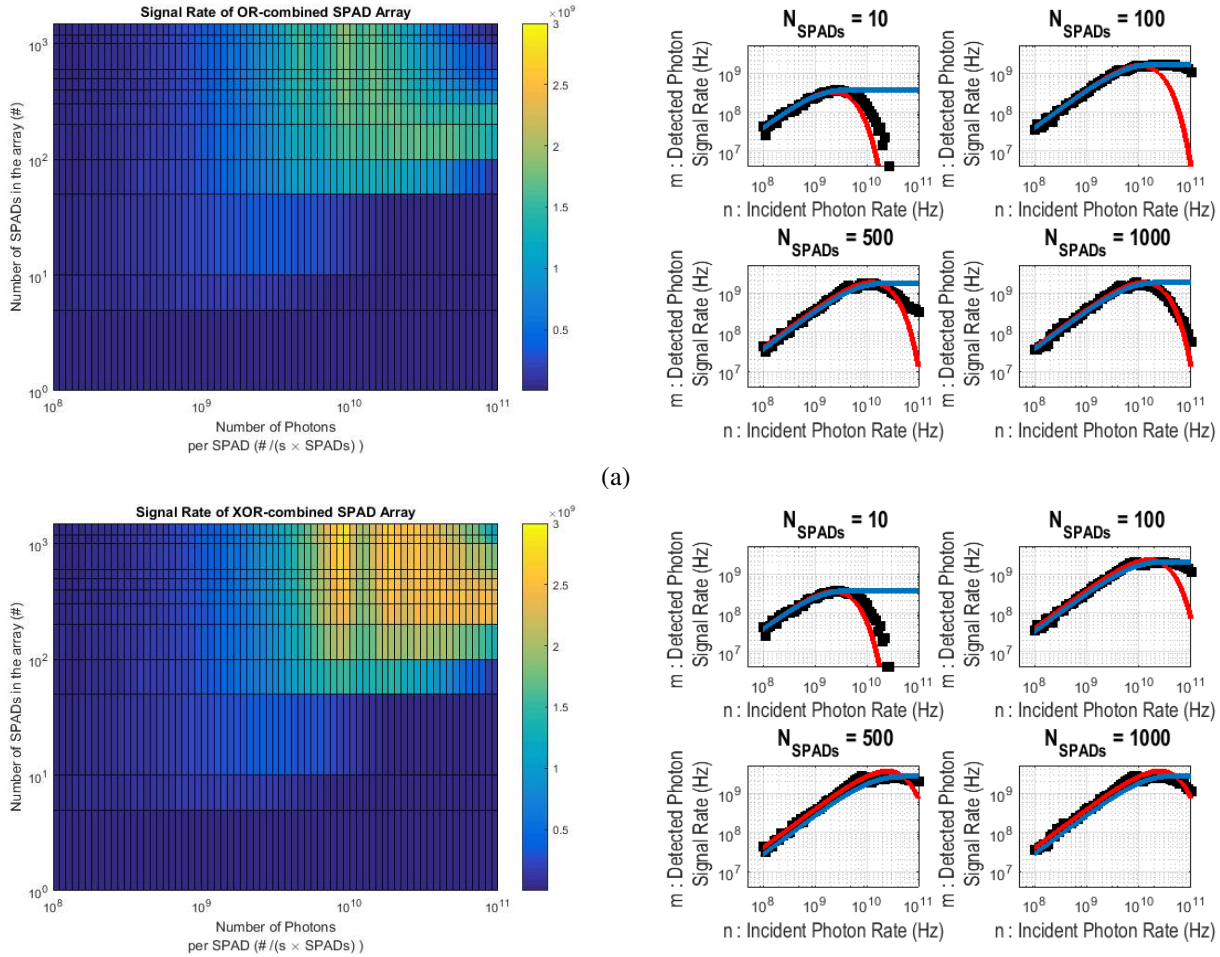


Figure 10. **Signal rate analysis results for $\sim 100\%$ fill factor array** - (a) OR-tree array with $\tau_{PS} + \tau_{electr} = 100$ps (b) XOR-tree array with $\tau_{PW,min} = 100$ps. With high fill factor, the only limitation to the signal rate is the effective dead time of the array. All the data are compared with paralysis curve (in red) and non-paralysis curve (in blue).

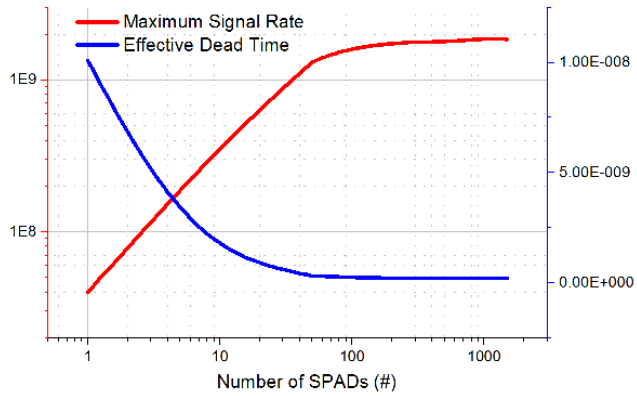


Figure 11. **Effective dead time** - The predicted effective dead time given by (14) is compared to the point of maximum signal rate and the value of the maximum taken by the simulated data

B. Fixed SPAD Array Area

It is of major interest to understand how many SPADs could be efficiently fitted in a certain given area to maximise the signal rate. Contrarily to the previous assumption, when more SPADs are fitted within the same area A_{total}, each SPAD becomes smaller:

$$A_{\text{SPAD}} = \frac{A_{\text{total}}}{N_{\text{SPADs}}} \tag{17}$$

The main consequences of this scaling can be assumed as the following:

- **Reduction of fill factor:** assuming that the same logic is used in each SPAD pixel in an area of A_{logic}, the fill factor FF is reduced as:

$$FF = 1 - \frac{A_{\text{logic}}}{A_{\text{SPAD}}} = 1 - \frac{A_{\text{logic}} \times N_{\text{SPADs}}}{A_{\text{total}}} \tag{18}$$

- **Reduction of dead time:** the dead time is proportional to the timing constant of the RC quenching circuit. When the SPAD area is scaled down to fit more number of SPADs within the same area, the resistance R can be

considered constant whereas the capacitance C scales with the area. The dead time per single SPAD is therefore given by:

$$\tau_{\text{dead}} \propto RC \propto A_{\text{SPAD}} \propto \frac{1}{N_{\text{SPADs}}} \tag{19}$$

In addition, aggregating more SPADs also decrease the overall dead time of the array as $\propto N_{\text{SPADs}}$. The two effects combined together leads to:

$$\tau_{\text{dead,eff}} \propto \frac{1}{N_{\text{SPADs}}^2} \tag{20}$$

The reduction of the fill factor and the reduction of the overall dead time go against each other in the direction of increasing the signal rate: a lower dead time increases the signal rate according to (15) whereas a lower fill factor decreases the signal rate by reducing the quantity of collected photons due to the lower active area of the SPAD. The effect of the fill factor can be modelled in the equation (15) by simply replacing the photon rate n with the effective photon rate n':

$$n'(N_{\text{SPADs}}) = n \times FF = n\left(1 - \frac{A_{\text{logic}} \times N_{\text{SPADs}}}{A_{\text{total}}}\right) \tag{21}$$

Simulations are run to get the efficient number of SPADs, for a given logic area, that maximises the signal rate under such conditions. Fig. 12 shows the results for a simulation run by also changing the value of the $A_{\text{logic}}/A_{\text{SPAD}}$. When the logic size is sufficiently small, the active area of the $N_{\text{SPADs}} < N_{\text{SPADs, best}}$ is not affected by the presence of the logic and the signal rate reaches the maximum set by the technology (τ_{min}). When the single SPAD pixel is occupied by a large logic, the process of scaling down the size to fit more pixels has an important impact on the signal rate. As the last two curves of Fig. 12 show, the fill factor goes down to zero far before reaching the $N_{\text{SPADs, best}}$. In these cases, the theoretical maximum signal rate set by the technology cannot be achieved: the fill factor is the predominant factor limiting the photon count rate.

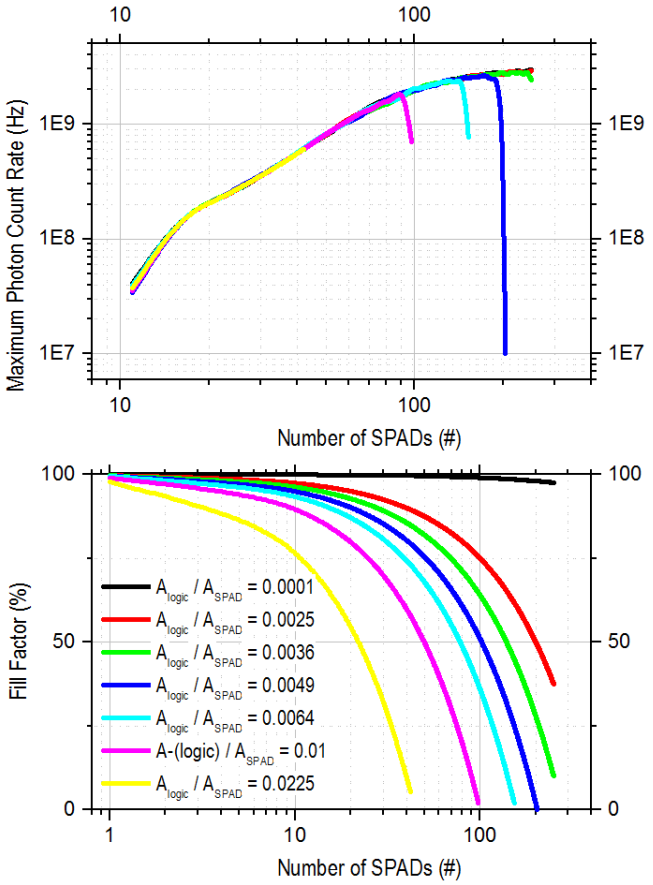


Figure 12. **Maximum signal rate** - The presence of logic inside the pixel influences the fill factor and therefore the maximum signal rate achievable in the process of scaling down the SPAD pixel in the fixed total area

V. Conclusions

The simulator offers an efficient way to test and model the behaviour of digital SiPMs in terms of photon count rate under a variety of assumptions: pixel area, number of SPADs, fill factor and combining logic. The effectiveness of the model can be improved by the addition of: dark count, after-pulse, cross-talk and the effect of ambient light. The algorithms have been implemented in MATLAB. The utility of the model is to guide future digital SiPM design to confirm or improve the model itself. A variety of applications where the signal rate is crucial, such as close range proximity sensors, will benefit from the results found by the simulator.

References

[1] F. Zappa, M. Ghioni, S. Cova, L. Varisco, B. Sinnis, A. Morrison, and A. Mathewson, "Integrated array of avalanche photodiodes for single-photon counting," in *Solid-State Device Research Conference, 1997. Proceeding of the 27th European*, September 1997, pp. 600–603.

[2] C. Piemonte, R. Battiston, M. Boscardin, G.-F. Dalla Betta, A. Del Guerra, N. Dinu, A. Pozza, and N. Zorzi, "Characterization of the First Prototypes of Silicon Photomultiplier Fabricated at ITC-irst," *Nuclear Science, IEEE Transactions on*, vol. 54, no. 1, pp. 236–244, Feb 2007.

[3] P. Buzhan, B. Dolgoshein, A. Ilyin, V. Kantserov, V. Kaplin, A. Karakash, A. Pleshko, E. Popova, S. Smirnov, Y. Volkov, L. Filatov, S. Klemin, and F. Kayumov, "An Advanced Study of Silicon Photomultiplier," *ICFA Instrum.Bull.*, vol. 23, pp. 28–41, 2001.

[4] T. Frach, G. Prescher, C. Degenhardt, R. de Gruyter, A. Schmitz, and R. Ballizany, "The Digital Silicon Photomultiplier - Principle of Operation and Intrinsic Detector Performance," in *Nuclear Science Symposium Conference Record (NSS/MIC), 2009 IEEE*, Oct 2009, pp. 1959–1965.

[5] Philips Digital Photon Counting. [Online]. Available: http://www.digitalphotoncounting.com/

[6] D. R. Schaart, E. Charbon, T. Frach, and V. Schulz, "Advances in digital sipms and their application in biomedical imaging," *Nucl. Instrum. Meth. A*, vol. 809, pp. 31 – 52, 2016, advances in detectors and applications for medicine.

[7] C. Niclass, M. Soga, H. Matsubara, S. Kato, and M. Kagami, "A 100m-Range 10-Frame/s 340 × 96-Pixel Time-of-Flight Depth Sensor in 0.18- μm CMOS," *IEEE J. Solid-State Circuits*, vol. 48, no. 2, pp. 559–572, Feb 2013.

[8] D. Tyndall, B. R. Rae, D. D.-U. Li, J. Arlt, A. Johnston, J. a. Richardson, and R. K. Henderson, "A high-throughput Time-Resolved Mini-Silicon Photomultiplier with Embedded Fluorescence Lifetime Estimation in 0.13 μm CMOS," *IEEE Trans. Biomed. Circuits Syst.*, vol. 6, no. 6, pp. 562–70, Dec. 2012.

[9] E. Webster, R. Walker, R. Henderson, and L. Grant, "A silicon photomultiplier with $> 30\%$ detection efficiency from $450 - 750$nm

and 11.6 μm pitch NMOS-only pixel with 21.6% fill factor in 130nm CMOS," in *Solid-State Device Research Conference (ESSDERC), 2012 Proceedings of the European*, Sept 2012, pp. 238–241.

[10] L. H. C. Braga, L. Gasparini, L. Grant, R. Henderson, N. Massari, M. Perenzoni, D. Stoppa, and R. Walker, "A Fully Digital 8×16 SiPM Array for PET Applications With Per-Pixel TDCs and Real-Time Energy Output," *IEEE J. Solid-State Circuits*, vol. 49, no. 1, pp. 301–314, 2014.

[11] N. Dutton, S. Gnecchi, L. Parmesan, A. Holmes, B. Rae, L. Grant, and R. Henderson, "A Time-Correlated Single-Photon-Counting Sensor with 14GS/s Histogramming Time-to-Digital Converter," *International Solid State Circuit Conference - Proceedings Of*, 2015.

[12] J. Arlt, D. Tyndall, B. R. Rae, D. D.-U. Li, J. a. Richardson, and R. K. Henderson, "A Study of Pile-up in Integrated Time-Correlated Single Photon Counting Systems," *Review of Scientific Instruments*, vol. 84, no. 10, pp. 103 105–103 105–10, Oct. 2013.

[13] Q. Li and Q. Hu, "A 10ps 500MS/s two-channel Vernier TDC in 0.18um CMOS technology," in *Advanced Research and Technology in Industry Applications (WARTIA), 2014 IEEE Workshop on*, Sept 2014, pp. 1268–1271.

[14] S. Cova, M. Ghioni, A. Lacaita, C. Samori, and F. Zappa, "Avalanche Photodiodes and Quenching Circuits for Single-Photon Detection," *Appl. Opt.*, vol. 35, pp. 1956–1976, 1996.

[15] G. Giustolisi, R. Mita, and G. Palumbo, "Verilog-A modeling of SPAD statistical phenomena," in *Circuits and Systems (ISCAS), 2011 IEEE International Symposium on*, May 2011, pp. 773–776.

[16] R. Mita, G. Palumbo, and P. Fallica, "Accurate Model for Single-Photon Avalanche Diodes," *Circuits, Devices Systems, IET*, vol. 2, no. 2, pp. 207–212, April 2008.

[17] D. Marano, M. Belluso, G. Bonanno, S. Billotta, A. Grillo, S. Garozzo, G. Romeo, O. Catalano, G. La Rosa, G. Sottile, D. Impiombato, and S. Giarrusso, "Silicon Photomultipliers Electrical Model Extensive Analytical Analysis," *Nuclear Science, IEEE Transactions on*, vol. 61, no. 1, pp. 23–34, Feb 2014.

[18] G. Giustolisi, P. Finocchiaro, A. Pappalardo, and G. Palumbo, "Behavioral Model of Silicon Photo-Multipliers suitable for Transistor-level Circuit Simulation," *Nuclear Science, IEEE Transactions on - in review*, 2014.

[19] S. Seifert, H. van Dam, J. Huizenga, R. Vinke, P. Dendooven, H. Lohner, and D. Schaart, "Simulation of Silicon Photomultiplier Signals," *Nuclear Science, IEEE Transactions on*, vol. 56, no. 6, pp. 3726–3733, Dec 2009.

[20] H. van Dam, S. Seifert, R. Vinke, P. Dendooven, H. Lohner, F. Beekman, and D. Schaart, "A Comprehensive Model of the Response of Silicon Photomultipliers," *Nuclear Science, IEEE Transactions on*, vol. 57, no. 4, pp. 2254–2266, Aug 2010.

[21] A. Eisele and R. Henderson, "185 MHz Count Rate, 139 dB Dynamic Range Single-Photon Avalanche Diode with Active Quenching Circuit in 130nm CMOS Technology," *Int. Image Sensor Workshop (IISW)*, pp. 6–8, 2011.

[22] V. Savuskan, I. Brouk, M. Javitt, and Y. Nemirovsky, "An Estimation of Single Photon Avalanche Diode (SPAD) Photon Detection Efficiency (PDE) Nonuniformity," *Sensors Journal, IEEE*, vol. 13, no. 5, pp. 1637–1640, May 2013.

[23] W. Leo, *Techniques for Nuclear and Particle Physics Experiments*. Springer, 1994.