Analysis of Photon Detection Efficiency and Dynamic Range in SPAD based Visible Light Receivers

Citation for published version:

Digital Object Identifier (DOI):
10.1109/JLT.2016.2550497

Link:
Link to publication record in Edinburgh Research Explorer

Document Version:
Peer reviewed version

Published In:
Journal of Lightwave Technology

General rights
Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy
The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.
Analysis of Photon Detection Efficiency and Dynamic Range in SPAD based Visible Light Receivers

*ST Microelectronics Imaging Division, Edinburgh, United Kingdom
†The University of Edinburgh, Edinburgh, United Kingdom

Abstract—We investigate the photon detection efficiency and the dynamic range for digital silicon photomultipliers (dSiPMs) over a selection of design parameters: dSiPM unit cell dead time, photon detection efficiency, unit cell area and fill factor, number of cells and total dSiPM active area. Two receiver scaling scenarios are considered: varying the number of cells for (1) a fixed unit cell area or (2) a fixed total dSiPM area. Theoretical and simulated results are confirmed with experimental data from a selection of dSiPMs realised on a test chip in 130nm CMOS process.

Index Terms—Single Photon Avalanche Diodes, SPAD, dSiPM, digital Silicon Photomultiplier, Visible Light Communication, VLC

I. INTRODUCTION

CMOS single photon avalanche diodes (SPADs) are finding commercial application in positron emission tomography (PET), time-of-flight ranging and advanced microscopy thanks to their high timing resolution, integration in array formats with fast digital signal processing at low cost [1], [2], [3], [4], [5], [6].

Recent interest has been shown in applying these detectors in guided wave or free space visible light communications (VLC) where they promise high sensitivity and photon shot noise limited links [7], [8]. Although only modest data rates have been so far obtained [9], [10], evidence of optical communications operating at the quantum limit is already emerging [11], [12], [13], [14]. Enhancements of data rates towards the Gb/s rates competitive with state of the art visible communications, [15], [16], [17], [18], rely on advances in the architectures of SPAD receivers. Two keys areas are being investigated (1) architectures for combining multiple SPAD outputs into a single sampled data stream (2) optimisation of the physical characteristics of the SPAD array itself. Analog circuit approaches to the latter problem include commercial analogue silicon photomultipliers (SiPMs) [19] as well as custom CMOS active analogue SiPM achieving 200Mb/s [20]. Digital silicon photomultipliers (dSiPMs) [21] first proposed for PET have also been applied to VLC providing direct integration of a “light to digital” electronic receiver with advantages of low power, circuit area and compatibility with existing DSP.

In this paper, we investigate theoretically and experimentally the trade-offs in selection of the number, dead time, fill-factor and area of unit cells in a dSiPM receiver approach in order to achieve certain sensitivity, dynamic range (DR), linearity and signal to noise ratio (SNR) properties. The results are relevant to future analogue and digital dSiPM VLC receiver architectures, particularly the case of bandwidth limited links such as found in free-space GaN LED [22] or polymer optical fibre (POF) [23]. In these cases, the potential SNR improvement offered by dSiPM receivers over APD or PIN solutions [8] requires use of higher order of modulation schemes (e.g. OFDM, PAM [10], [13]) and linear transmitter and receiver characteristics.

Our study focusses on a recently proposed technique to combine multiple cell outputs into a data stream; the XOR tree (Fig. 1). This approach has been shown to be effective in recent PET dSiPMs and proof-of-concept VLC receivers [11], [24]. A series of experimental XOR dSiPMs has been constructed in 130nm CMOS test chip allowing the number, diameter, dead time and dark count rate (DCR) of the SPAD cells in XOR dSiPM receiver front-ends to be varied. Measurements of important properties for communication system designers are provided allowing linearity, DR, SNR and photon detection efficiency (PDE) (or sensitivity) to be directly derived. In addition, non-linearity and saturation limits are studied allowing practical requirements to be set on received signal power.

Although an XOR dSiPM has been used as the basis of this study, many results are valid for analogue SiPMs and other pulse-combining readouts, [21], [25], [26].

The paper examines two dSiPM receiver design scenarios (1) a chosen fixed unit cell area with no limitation on total area (hence number of cells) combined by one XOR tree or (2) a fixed total dSiPM area with the possibility of fitting more (but smaller) cells in the chosen area at the cost of fill-factor and sensitivity. Section I and
II. DIGITAL SiPM WITH FIXED UNIT CELL AREA

To realise high count rate detectors, arrays of SPAD cells are manufactured together with timing or counting circuits in so-called silicon photomultipliers (SiPMs) [27], [28], [29]. As suggested by Mandal et al. [2], the structure of a SiPM can be optimised by designing arrays with outputs combined into a common pulse-combining readout channel. This technique allows timing or counting circuitry to be shared thus reducing the required silicon area and simplifying the read-out. The many advantages of digital aggregation of SPAD cells has led to the increasing popularity of digital silicon photon multipliers (dSiPMs) [31]. A typical way to digitally combine SPAD cells is by the use of an OR-tree preceded by a monostable pulse-shortener cell per cell [23], [26]. The limitation on the count rate due to the latter has been overcome by the replacement of monostables and OR tree respectively with toggle cells and XOR tree [11], [24], see Fig. 1 promising high count rates in optimised dSiPMs. To investigate the optimisation process, we analyse the performance of such detectors in terms of photon detection efficiency and dynamic range. Together with the estimation of such figures of merit under a selection of assumptions, we provide a comparison with experimental data to verify the effectiveness of the two approaches.

One typical option in dSiPM design is to aggregate a certain number of unit cells with the same active area $A_{cell}$ and dead time $\tau_d$. It is of interest to understand how the dynamic range changes when different number cells are aggregated into a common pulse-combining readout receiver.

A. SiPM Photon Detection Efficiency

For an individual dSiPM unit cell, the photon detection efficiency ($PDE_1$) is calculated as the ratio between the count rate $m$ and the rate $n$ of the incident photons on the total area of the SPAD cell including any per-SPAD circuitry such as guard rings, well isolation, quench circuits, memory, buffering or pulse combining electronics. The fill-factor ($FF$) is the ratio of photosensitive area to SPAD unit cell area. For an SiPM made of identical cells, the total $PDE$ is the same as a single cell since both the count rate $m_{TOT}$, assuming no loss of counts (low light level), and the incident photons $n(N)$ scale linearly with the number of diodes:

$$PDE(N) = \frac{m_{TOT}}{n(N)} = PDE_1$$  \hspace{1cm} (1)

While at moderate light level the total count rate of a dSiPM is proportional to the incident photon rate (dSiPM $\lambda_{total}$ is fixed $A_{cell}$)

$$\lambda_{total} = N \times A_{cell}$$

(a) $A_{total} = \lambda_{cell}$

(b) $A_{total} = 4\lambda_{cell}$

(c) $A_{total} = 16\lambda_{cell}$

Figure 2. Fixed unit cell area for dSiPM design - A different number of cells with fixed area can be integrated into arrays. The total area of the dSiPM scales linearly with the number of cells: $\lambda_{total} = N \times \lambda_{cell}$.

linear region), at high light levels, due to detector saturation a loss of registered counts is observed. This region will be referred to as saturation region.

B. Dynamic Range

We define as dynamic range $DR$ the ratio between the maximum registered count rate and the noise level known as the dark count rate ($DCR$) of the dSiPM:

$$DR (dB) = 20 \cdot \log_{10} \frac{m_{MAX}}{DCR}$$  \hspace{1cm} (2)

For a dSiPM made of $N$ identical unit cells, the $DCR$ is linear with the number of cells, hence we write:

$$DCR(N) = N \cdot DCR_1$$  \hspace{1cm} (3)

At high incident photon rates, a dSiPM enters its saturation regime where, due to count loss, the count rate is not proportional to the incident photon rate. The count loss can be caused by either the saturation of the single diodes, or the bandwidth limitation imposed by the pulse-combining readout channel. The former is described by the dSiPM unit cell dead time $\tau_d$ while the latter can be described by an equivalent maximum frequency $f_{BW}$ limiting the recorded count rate. The maximum count rate $m_{MAX}$ for a dSiPM is modelled as:

$$m_{MAX}(N) = f_{BW} \left( 1 - e^{-N \cdot m_{1}/f_{BW}} \right)$$  \hspace{1cm} (4)

where $m_{1}$ is the maximum detection rate of an individual cell equal to $1/(e \cdot \tau_d)$ for a passive recharge SPAD cell [30]. This expression fits experimental data, see Section IV with a level of confidence described by a reduced chi-squared $\chi^2 \sim 0.9$.

We then substitute (2) and (3) in (4):

$$DR (dB) = 20 \cdot \log_{10} \left( \frac{1 - e^{-N/(e \cdot \tau_d \cdot f_{BW})}}{N \cdot DCR_1} \right)$$  \hspace{1cm} (5)

For $N \ll e \cdot \tau_d \cdot f_{BW}$ the maximum count rate is:

$$m_{MAX}(N) \sim \frac{N}{e \cdot \tau_d}$$  \hspace{1cm} (6)
A. SiPM Photon Detection Efficiency

To calculate the photon detection efficiency of a dSiPM under such an assumption, we need to consider one important effect of shrinking the dSiPM unit cells: the decreasing fill factor due to the in-pixel electronics. When \( N \) cells are fabricated in the given area, each will have a fill factor of \( FF(N) \), therefore:

\[
PDE(N) = PDE_1 \cdot \bar{FF}(N)
\]

(10)

where \( PDE_1 \) is the PDE of the detector when only one cell (of \( FF_1 \) fill factor) is fitted in the available area, and \( \bar{FF}(N) \) is the reduction of fill factor when the area is populated by \( N \) cells:

\[
\bar{FF}(N) = \frac{FF(N)}{FF_1}
\]

(11)

Such term is highly dependent on the dSiPM unit cell design and will not be modelled in this work and will be left as a factor to be plugged in for final calculations.

B. Dynamic Range

We now calculate the DCR and the maximum signal rate of the dSiPM for a fixed total area. It is well known that the DCR scales with the active area of the dSiPM \([31]\), therefore in the scenario of \( N \) cells fitted into a constant total area, we can approximate the DCR over the dSiPM as:

\[
DCR(N) = DCR_1 \cdot \bar{FF}(N)
\]

(12)

where \( DCR_1 \) is the DCR of the device when only one cell is fitted into the total available area.

For the calculation of the maximum count rate, two aspects need to be considered: when smaller dSiPM unit cells are fabricated, they typically exhibit lower dead times compared to bigger cells due to the size of the device capacitance. Therefore the maximum count rate can be written as a variation of (4):

\[
m_{\text{MAX}}(N) = f_{BW} \left( 1 - e^{-N \cdot m_1(N)/f_{BW}} \right)
\]

(13)

where \( m_1(N) = 1/(e \cdot \tau_d(N)) \). However, when the number \( N \) increases, the maximum count rate is no longer dominated by the individual dSiPM unit cell dead time \( \tau_d(N) \), instead it is limited by the term \( f_{BW} \) describing the bandwidth of the pulse-combining readout. In this work, we concentrate on this behaviour rather than well known saturation due to dead time. Therefore, for simplicity, let us consider the variation on the dead time as second order effect so that we can replace the term \( \tau_d(N) \) with a constant \( \tau_d \) and use (4) instead of the more generic (13).

We can now estimate the dynamic range as:

\[
DR(\text{dB}) = 20 \cdot \log_{10} \frac{f_{BW} \left( 1 - e^{-N \cdot m_1(N)/f_{BW}} \right)}{DCR_1 \cdot \bar{FF}(N)}
\]

(14)
Dynamic Range and Photon Detection Efficiency - Definition and Theoretical Expressions Under the Two Scenarios of Sections II and III

<table>
<thead>
<tr>
<th>Definition</th>
<th>Fixed Total dSiPM Area</th>
<th>Fixed dSiPM Area</th>
</tr>
</thead>
<tbody>
<tr>
<td>Linear Dynamic Range</td>
<td>( DR(dB) = 20 \cdot \log_{10} \frac{N}{e \cdot \tau_d \cdot DCR_1 \cdot FF(N)} ) for ( N \ll e \cdot \tau_d \cdot f_{BW} )</td>
<td>( DR(dB) = 20 \cdot \log_{10} \frac{f_{BW}}{DCR_1 \cdot FF(N)} ) for ( N \gg e \cdot \tau_d \cdot f_{BW} )</td>
</tr>
<tr>
<td>Photon Detection Efficiency</td>
<td>( PDE(N) = \frac{N_{TOT}}{n} )</td>
<td>( PDE_1 = PDE ), ( \tilde{PDE}(N) = PDE_1 \cdot \tilde{FF}(N) )</td>
</tr>
</tbody>
</table>

The two limits regarding the number of unit cells and the minimum dead time are here presented:

\[
DR(dB) \approx 20 \cdot \log_{10} \frac{N}{e \cdot \tau_d \cdot DCR_1 \cdot FF(N)} , \text{ for } N \ll e \cdot \tau_d \cdot f_{BW} \tag{15}
\]

\[
DR(dB) = 20 \cdot \log_{10} \frac{f_{BW}}{DCR_1 \cdot FF(N)} , \text{ for } N \gg e \cdot \tau_d \cdot f_{BW} \tag{16}
\]

Results from both analyses are summarised in Table I and will be compared to experimental data in the next section.

IV. Test Chip

We present now a test chip manufactured in STMicroelectronics 130nm imaging process as shown in Fig. 4. Two sets of dSiPMs of previously published SPAD structures have been designed, [31], [32]. The crosstalk, less than 1%, has not been included in the analysis since considered a second order effect. The first set is composed by a 16 x 16 XOR-combined with 7μm pitch shared-well passively quenched SPAD cells. The second set consists of five pitch variants of 4 x 4 arrays of the same structure, with external combination logic. The 16 cell outputs from each dSiPM are multiplexed onto selectable XOR and OR trees. Table II summarises the properties of the designed dSiPM. In our test dSiPMs, the unit cells contain only SPADs wired to readout electronics placed at the exterior of the arrays. This has been done to avoid interactions between SPADs and neighbouring electronics. The fill-factor is therefore higher than practically achievable in larger dSiPM arrays but there is no loss of generality in our model. A 16bit on-chip ripple counter provides the counts \( M \) of the selected dSiPM (and relative enabled cells) for a controlled exposure time \( T_{exp} \) from which the average count rate \( m \) is calculated as \( m = M/T_{exp} \).

The SPAD cells were biased to the same average dead time of \( \tau_d = 5ns \) in order to fulfil the assumption useful for calculations. The bandwidth of the pulse-combining readout consisting of the XOR tree and the on-chip counters is measured as \( f_{BW} = 1.01GHz \).

For the experiments, we illuminate the detector with an LED with dominant wavelength \( \lambda = 470nm \). For each light intensity we record the average count rate. The exposure time has been set to a range from 2μs to 0.6ms in order to have significant number of counts in the 8bit on-chip ripple counter and to avoid its saturation over all the desired light levels. To improve the statistics, each exposure is repeated for 500 iterations taking advantage of the off-chip memory (PC) to store larger and therefore more statistically significant data. For all the obtained count rates, we report the mean value and the standard deviation in error bar plots.

A. dSiPM D1

We first select the dSiPM D1 to perform light intensity sweeps for an increasing number of activated cells from 1 to the maximum 256 available on chip. This mimics the assumptions of Section II.

We start by enabling an increasing number of activated cells. We show three cases in Fig. 5. With only one cell activated (purple line), the dSiPM shows a low DCR level but a limited maximum count rate. The middle green line represents the intermediate case of number of cells \( N = 16 \). In this configuration, the maximum count rate has significantly risen and the dynamic range is not affected. The extreme case of all 256 cells activated, shown as a light blue line, confirms the model employed in Section

<table>
<thead>
<tr>
<th>dSiPM</th>
<th>Unit Cell Pitch (μm)</th>
<th>Number of Cells</th>
<th>Fill Factor (%)</th>
<th>Dead Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>D1</td>
<td>7 x 16</td>
<td>16 x 16</td>
<td>6.4</td>
<td>54ns</td>
</tr>
<tr>
<td>D2</td>
<td>9 x 4</td>
<td>4 x 4</td>
<td>18.7</td>
<td>37.4</td>
</tr>
<tr>
<td>D3</td>
<td>13 x 4</td>
<td>4 x 4</td>
<td>73.6</td>
<td>54ns</td>
</tr>
<tr>
<td>D4</td>
<td>18.62 x 4</td>
<td>4 x 4</td>
<td>85.4</td>
<td></td>
</tr>
<tr>
<td>D5</td>
<td>34.62 x 4</td>
<td>4 x 4</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 4. Test Chip - A selected samples of dSiPMs has been manufactured on the same test chip. All the varieties of dSiPMs have been combined through toggle cells and a common XOR tree outside the cells.
aggregating too many unit cells does not increase the maximum count rate, due to detector saturation, but highly affects the dynamic range by increasing the DCR level. Fig. 6 shows the measured dynamic range over a different number of enabled cells. The deviation from the model has to be ascribed to the non-uniformity of the dSiPMs which is assumed in the modelling. However, the closeness of the error bar to the predicted line confirms (9). The experiment shows the importance of keeping the number of aggregated cells low enough in order to avoid the pulse-combining readout channel saturation. In large dSiPM design, this implies that a multi-channel approach is more efficient in terms of higher dynamic range, such as the use of an array of mini-dSiPMs, [2], [26].

**B. dSiPM D2-D5**

The second set of dSiPMs (D2-D5) allows the assumptions of Section III to be experimentally validated. We choose the area occupied by a single cell of the dSiPM D5 in Table II as the reference total area. In the chosen area, one can then fit:

- 1 cell from D5
- 3 cells from D4
- 7 cells from D3
- 14 cells from D2

Therefore, we perform the light intensity sweeps choosing the desired dSiPM and the number of activated cells. The photon transfer curve for each configuration is shown in Fig. 7. Moreover, we provide direct measurements of the photon detection efficiency and the dynamic range in Fig. 8 to compare the data to the proposed equation model, respectively (10) and (14). As expected, having a larger number of smaller cells highly increases the DR, due to lower DCRs and higher maximum count rates, while the photon detection efficiency is negatively affected by the reduction of fill factor.

**V. NOISE AND SIGNAL-TO-NOISE RATIO**

From the obtained data, additional considerations on noise can be derived. We show in Fig. 9(a) the noise as the standard deviation of the measured counts (data points) compared directly with the photon shot noise calculated as the square root of the mean counts (solid line) both normalised for a unit exposure time. These experimental data demonstrate detection rates are limited only by photon shot noise in linear regime of operation, i.e. in the region where the count rate is proportional to the incident photon rate, essential for optical communications.

Moreover, we calculate the signal-to-noise ratio (SNR) for all the light intensity points as:

\[
\text{SNR} = 20 \log_{10} \frac{M}{\sigma_M}
\]

(17)

Results are shown in Fig. 9(b) where the mean value and the standard deviation of the counts have been again normalised for a unit exposure time. We confirm that the SNR increases linearly with the square root of the number of collected photons before the count loss due to the dSiPM saturation. Although smaller unit cells (purple line) show lower SNR compared to larger ones (yellow...
Figure 8. **Fixed total area** - The change in the photon detection efficiency (a) and dynamic range (b) is measured for different number of unit cells fitted in a total area. The graph shows the effect of fewer collected photons, reaching a higher maximum due to a higher saturation threshold. The graph also shows the noise floor region: below a certain light level, the constant count rate measured from the chip is pure DCR and does not contain signal information.

**VI. CONSIDERATIONS ON GENERAL DSiPM DESIGNS**

The modelling equations here proposed have been derived under a selection of assumptions. We now discuss how to apply necessary modifications in more general SiPM designs.

a) **Uniformity of the dSiPM:** one of the first assumptions made in the modelling is that the dSiPM consists of identical unit cells, therefore the dSiPM is supposed to have a strong uniformity in terms of fill factor and dark count rate. While the former is generally true, the latter is typically not verified especially in large dSiPMs. When more cells are enabled in a large dSiPM (case described in Section II) the calculation of the total DCR proposed in (3) can be replaced by the more general:

\[
DCR(N) = \sum_{i=1}^{N} DCR_i \quad (18)
\]

which will then replace the denominator of (5). When aggregating smaller and smaller cells, the approximation proposed by (12) might not be satisfied and therefore the general term \( DCR(N) \) should replace the denominator of (14). Similarly, if smaller cells show significant lower dead times, the assumption of \( \tau_d(N) = \tau_d \) can be dropped and the general expression of the maximum count rate (13) should be used in the dynamic range calculation. As previously stated, the dependency on the dead time becomes negligible in large pulse-combining dSiPMs where the main limitation to the dynamic range is given by the common readout bandwidth.

b) **Interconnection/Readout impact:** the maximum signal rate of the dSiPM has been shown to depend on both the dead time of the unit cells and on the pulse-combining readout bandwidth. The latter dependency becomes dominant when a large number of cells is combined onto a common readout. It is therefore useful during the design process to estimate the bandwidth of the pulse-combining readout taking into account also digital circuit switching speeds and interconnect parasitics. This will allow an estimation of the parameter \( f_{BW} \), present in all the proposed equations, and therefore of the dynamic range.

c) **In-Pixel Electronics:** the model here proposed has been verified with a test chip containing small in-pixel electronics (only buffers and toggle cells) while the main XOR tree is shared outside the cell arrays.

Figure 9. **Signal-to-Noise** - For the set of dSiPMs shown in Fig. 7 (a) the measured noise (data points) are plotted together with the ideal photon shot noise limit (solid line) and (b) the normalised signal over the standard deviation data are plotted for the whole range of light levels.
However, typical dSiPMs contain additional cells, such as monostable circuits and OR cells. The impact of such design choice on the model is the following: first of all, the fill factor of the dSiPM unit cells decreases due to the occupying cells. Therefore the parameter \( F \) present in the equations needs to take into the account such reduction. Moreover, having in-pixel electronics might impact the single-channel bandwidth, which yet again can be simulated and included in the model by estimating the proposed parameter \( f_{\text{FWHM}} \). With these parameters, the model allows general conclusions to be derived as well as the example test chip proposed in this work.

VII. CONCLUSIONS

We have calculated the photon detection efficiency and dynamic range of digital silicon photomultipliers depending on typical design parameters.

We have demonstrated that a high dynamic range, exceeding 110dB, can be obtained by using an array of small SPAD cells, \( 7 \mu m \) pitch, with an optimal number of activated cells \( \sim 16 \). We have moreover proven single photon shot noise count rates in the linear region of operation of the dSiPM. All these results can be applied to the modelling and design of future dSiPM receiver architectures.

REFERENCES

[3] C. Niclass, M. Soga, H. Matsuura, and M. Kagami, “A 0.18-\mu m CMOS SoC for a 100-