A 607 MHz time-compressive computational pseudo-dToF CMOS image sensor

Pham Ngoc Anh[1], Thoriq Ibrahim[1], Keita Yasutomi[2], Shoji Kawahito[2],
Hajime Nagahara[3], Keiichiro Kagawa[2]

[1] Graduate School of Integrated Science and Technology, Shizuoka University

3-5-1 Johoku, Naka Ward, Hamamatsu, Shizuoka, Japan 432-8011

[2] Research Institute of Electronics, Shizuoka University

3-5-1 Johoku, Naka Ward, Hamamatsu, Shizuoka, Japan 432-8011

[3] Institute for Datability Science, Osaka University

2-8, Yamadaoka, Suita, Osaka, Japan 565-0871

Email: kagawa@idl.rie.shizuoka.ac.jp

Abstract - **This paper demonstrates a pseudo-direct time-of-flight (pseudo-dToF) CMOS image sensor that is robust to multipath interference (MPI) and has high distance accuracy and precision. This method uses an iToF-based image sensor, but can reconstruct the received light waveforms similar to those obtained by conventional dToF image sensors based on single-photon avalanche diode (SPAD). Therefore, this method has the advantages of both dToF and iToF depth image sensors such as high resolution, high accuracy, immunity to MPI, and motion-artifact-free. This paper presents a signal reconstruction scheme for our laboratory-designed time-compressive image sensor based on the charge domain compressive sensing. Two approaches to refine the depth resolution are explained: 1) Increasing the operating clock speed; 2) Oversampling in image reconstruction and quadratic fitting in depth calculation. Experimental results show the separation of two reflections 40 cm apart under an MPI condition, and a significant improvement in distance precision down to 1 cm order. These results suggest that this method could be a promising approach to virtually implement dToF imaging suitable for challenging environments with MPI.**

I. INTRODUCTION

Time-of-flight (ToF) depth imaging calculates the distance of an object by measuring the travel time of light emitted from the camera to the object and back to the camera. This method is increasingly utilized in the fields of robotics and automotive applications. In ToF cameras, each pixel provides a specific depth value. However, the target scene may involve multiple light paths that interact with the same pixel, resulting in depth images that can contain scene-dependent errors due to multipath interference (MPI).

There are two major methodologies for traditional ToF depth imaging: direct ToF (dToF) and indirect ToF (iToF). The dToF sensor, based on the single electron avalanche diode (SPAD) [1], directly measures the reflected light waveforms, making it immune to MPI. However, it requires a larger circuit area for time-to-digital converters and histogram builders. In contrast, the iToF sensor [2] has a smaller circuit area and can estimate the depth with a higher spatial resolution, but it is susceptible to MPI. This is because it calculates the depth from the number of charges that correlate the incident light waveform with a demodulation function applied to the modulator.

In this work, we propose a new measurement method called pseudo-dToF, which provides the advantages of both dToF and iToF. This method utilizes iToF-based high-speed charge modulators, allowing for high-resolution imaging similar to iToF image sensors. Moreover, the time-compressive sensing in the charge domain enables the sensor to reproduce the entire light waveform in a single shot. Therefore, pseudo-dToF realizes high accuracy in depth and motion artifact-free measurement. We also present two approaches to achieve higher temporal resolution and report the improvements in depth accuracy and precision by experiments.
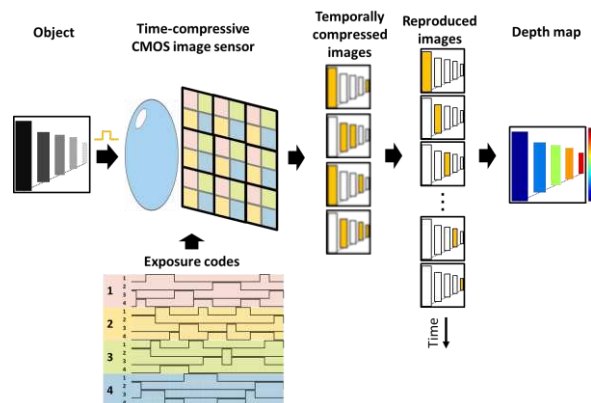
II. PSEUDO–DTOF DEPTH IMAGING

Fig. 1. Image acquisition and reconstruction flow

In our scheme, incident light signals are temporally compressed in the charge domain and reconstructed in three phases: pre-measurement, sensing, and signal reconstruction.

Firstly, we prepare exposure codes and measure the instrument response function (IRF) of the imaging optics and the image sensor in advance, which are used for signal reconstruction later.

In the sensing phase, the camera captures temporal signals of light emitted from a synchronized laser and reflected from objects. As shown in Fig. 1, the image sensor is composed of macro-pixels based on charge modulators. By applying exposure codes to the pixels

during the image shooting, multiple temporally compressed images (four in Fig. 1) are obtained at once.

In the reconstruction phase, the input optical waveforms for all subpixels are reproduced by solving the inverse problem based on the sparsity regularization. Thus, temporally sequential images or transient images of light are obtained. Subsequently, the temporal peak positions of the light waveform are detected with a quadratic curve fitting. Finally, they are converted to the object's depth using the speed of light.

a) Multi-tap macro-pixel CMOS image sensor

This sensor utilizes iToF-based charge modulators, which contribute to small pixel size or high-spatial-resolution imaging.
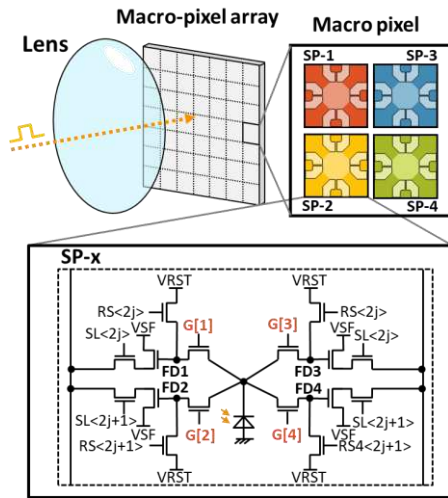

Fig. 2. Image sensor configuration.
One macro pixel is composed of 2×2 four-tap subpixels.

Figure 2 shows the pixel structure, which comprises an array of four sub-pixels (SPs) for each macro-pixel. Each sub-pixel is implemented by a four-tap LEFM charge modulator [3]. The LEFM charge modulator is composed of a photodiode and four sets of a charge transfer gate and a storage diode (tap). It outputs the integral value of a time-variant optical signal $g(t)$ within a designated time window function $\omega_i(t)$ as a pixel value Q_i (Eq. 1). Here, the index i identifies a tap.

$$Q_i = \int_0^{T_{exp}} g(t) \cdot \omega_i(t) dt \qquad \text{Eq. 1}$$

The high-speed operation of LEFM pixels allows for high temporal resolution. In our previous research, we confirmed that the sensor can be driven at a clock frequency of up to 303 MHz [4].

b) Compressive sensing

Compressive sensing [5] is an efficient sampling method that reconstructs more data points from fewer samples when the original signal is sparse. We employ this principle to reconstruct the entire incident light waveform from the compressed output images, result-ing in dToF-like signals with high-depth accuracy and robustness to MPI, like the conventional dToF method. Note that compressed images are obtained in a single shot, which makes our sensor motion-artifact-free.

Table 1. Sensor architecture

Technology	0.11 μm FSI CIS
Chip size	7.0 mm^H×9.3 mm^V
Macro-pixel size	22.4 μm^H×22.4 μm^V
Effective sub-pixel count	212^H×188^V
Sub-pixel count per macro-pixel	2^H×2^V
Maximum exposure code length	256 bits
Maximum modulation frequency	303 MHz
Maximum clock frequency	607 MHz
Image readout frame rate	21 fps
Power consumption	2.8 W

Our sensor utilizes a coded exposure pattern that is a temporal series of random binary values to create time-compressed signals. When the exposure code is 0, no charge is transferred to a specific tap, and when it is 1, the charge is stored in the tap. Therefore, the number of charges accumulated in each tap corresponds to the correlation value between the input light signal and the coded exposure pattern, resulting in temporally compressed images. The exposure codes are repeatedly applied to increase the pixel value.

Time-compressive sensing and signal reconstruction are as follows. Consider the input signal is x and the corresponding measurement signal is y, their relation can be expressed by the following linear equation:

$$y = Ax = (WH)Px \qquad \text{Eq. 2}$$

Here, matrix A is a spatio-temporal observation matrix that includes the spatial IRF of the imaging optics P, the exposure code W and the sensor's temporal IRF H. The dimensions of x, y, and A are N, M, and $M \times N$, respectively. In cases where $N > M$, i.e., the measured signal is lower in dimensionality than the original input signal, the signal is compressed. Retrieving the original input signal x from y is an ill-posed problem. However, if x is K-sparse, meaning that only K elements have non-zero values, we can determine the estimated solution for x by optimizing the L1 norm.

Based on this principle, the reflected light waveform x of all sub-pixels is reconstructed from the four compressed images y and the pre-measured observation matrix A. In this process, total variation is minimized instead of the L1 norm as shown in Eq. 3. Here, D_i shows a spatio-temporal differential operator. This process is performed by TVAL3 [6], a compressed sensing solver, using an iterative method.

$$\hat{x}^{(TV)} = \arg \min_x \sum_i \|D_i x\|_1, \text{ s.t. } y = Ax \qquad \text{Eq. 3}$$

Thus, transient images of light are reproduced, providing depth information. The depth accuracy and precision are dependent on the frame rate of the transient images. The temporal resolution of the reconstructed signal is determined by the reciprocal of

this frame rate, which is also equal to the minimum duration of the exposure code. For example, if our sensor operates at 303 MHz and uses a 32-bit exposure code, the temporal resolution is 3.3 ns (0.5 m in depth), and the measurable distance is 16 m.

To generate a depth map from the reconstructed images, the reconstructed waveform for each pixel is analyzed to determine the temporal peak position, which corresponds to the round-trip ToF. Finally, it is converted into a depth value using the speed of light.

III. IMPROVEMENT OF DEPTH RESOLUTION

So far, we have succeeded in ToF imaging at an operating clock frequency of 303 MHz [4]. This allowed the minimum time window with a duration of 3.3 ns, which corresponded to a 0.5 m resolution. Generally, higher depth resolution can be achieved simply by increasing the operating frequency of the charge modulator. However, driving the charge modulation at a higher frequency is highly dependent on the fabrication technology and modulator structure. It was challenging to achieve a minimum time window duration of less than 3.3 ns for our modulators. To overcome such restrictions, the following two approaches were applied to obtain higher depth resolution.

a) Sub-time-window shifting

Here, sub-time-window shifting of the exposure code is introduced to enhance the temporal resolution without changing the minimum time window duration, which is determined by the modulator design. To implement the sub-time-window shifting, we double the PLL clock frequency from 303 MHz to 607 MHz, while maintaining the minimum time window duration at 3.3 ns, equivalent to two clocks. Then, as shown in Figure 4, the falling and rising edges of the time window are shifted by one clock, which is half of the minimal time window. As a result, the temporal resolution is halved down to 1.75 ns.

Figure 5 shows the temporal sensor IRF measured

Fig. 5. Measured temporal sensor IRF
for the random codes at 607 MHz

b) Oversampling

Another approach to improve the temporal resolution is oversampling the sensor's IRFs, i.e., increasing the number of data points of the pre-measured observation matrix used for reconstruction. Previously, the sensor response was sampled once per bit of the exposure code. However, in this study, we have sub-sampled 10 points per bit (10× oversampling), as depicted in Fig. 6. The temporal (or depth) resolution is then improved, resulting in a denser reconstructed light waveform. Since the signal waveforms can be obtained as in dToF imaging, we can perform fitting to refine the depth. After the peak position in the waveform is detected from the reproduced signal waveform, the five data points around the peak are considered in fitting with a quadratic curve.
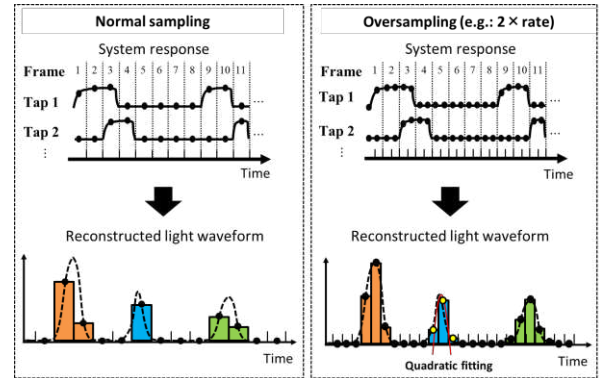
Fig. 6. Oversampling pre-measured sensor's IRF and fitting

The proposed method works better with oversampling, as denser light waveforms are utilized in fitting, resulting in more accurate depth estimation. Moreover, even if there are multiple peaks due to MPI, each peak can be distinguished and separately detected. Note that the depth resolution can be improved without changing the image sensor hardware or the total measurement time, although the reconstruction process takes longer.

IV. EXPERIMENTS

To evaluate the improvement of depth resolution after increasing the clock frequency from 303 MHz to 607 MHz, we compared the peaks separation performance under MPI conditions. In this experiment,
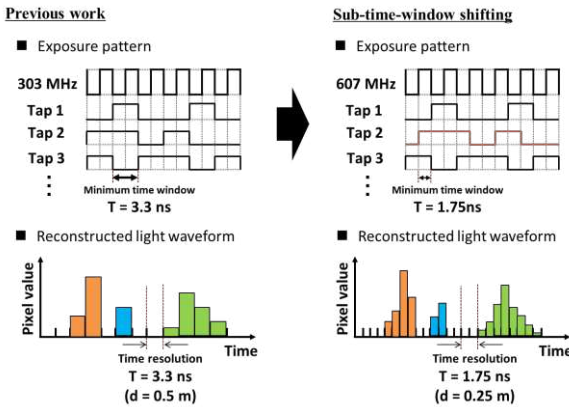

Fig. 4. Improvement of temporal resolution
by sub-time-window shifting

when this sensor was operated at 607 MHz. The exposure code was a 32-bit random binary code, and a 445 nm laser with an FWHM of 60 ps was used. The results were averaged for each tap. It can be observed that charge modulation was successfully performed.
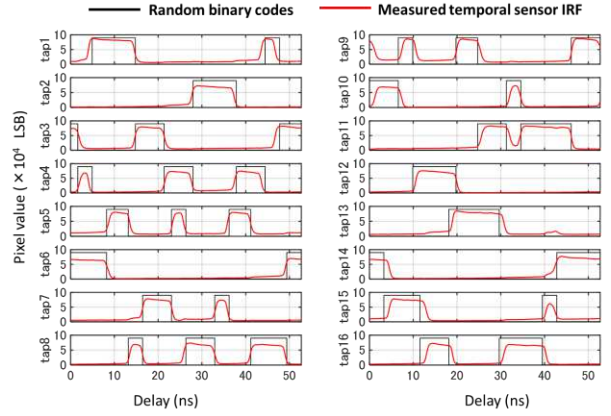
interference light was introduced by a weak diffusive plastic sheet placed in front of an objective panel with the letters "SU" as shown in Fig. 7a. The letters look blurry due to the diffuser. The distance between the two objects was set to 0.65 m and 0.40 m, and experiments were conducted with the sensor operating at 303 MHz and 607 MHz. A short-pulse semiconductor laser (PicoQuant LDH-IB-450-M-P, 443 nm) with a pulse width of 228 ps was used.

Figure 7b shows the reconstruction of optical waveforms at two clock frequencies. At 303 MHz, the interference light due to the diffuser was merged with the objective reflection for both clearances and could not be separated. However, when the clock frequency was increased to 607 MHz, the two reflections were completely separated for the 0.65 m clearance. For the 0.40 m clearance, although there was an overlap between the two signals, the object's reflection peak could be distinguished. These results show that the MPI is resolved in the 607 MHz operations.

Next, a ToF imaging experiment was conducted to verify the effectiveness of oversampling for a ratio of 10. Note that the time/depth resolutions at 303 MHz, 607 MHz, and 607 MHz with 10× sampling are 3.3 ns/0.5 m, 1.75 ns/ 0.25 m, and 0.175 ns/ 0.025 m respectively. A pulsed semiconductor laser with a wavelength of 660 nm and an FWHM of 2.5 ns was used. Figure 8 shows the configuration of the targets and the depth maps for each condition. 100-image averaging was applied to improve the SNR. Fig. 9 compares the reconstructed optical waveforms and their fitting curves with a quadratic approximation. The mean and standard deviation of the depth is quantitatively compared in Table 1, which demonstrates the effectiveness of the proposed method.

IV. CONCLUSION

In this paper, we demonstrated the concept, benefits, and implementation of pseudo-dToF imaging. Several techniques such as increasing clock frequency from 303 MHz to 607 MHz and oversampling have been applied to improve the depth resolution and estimated depth precision. While there are still some issues such as long processing time, vulnerability to ambient light and so on, the proposed method remains a promising technique for applications like autonomous vehicles, and robotics in challenging environments with MPI.
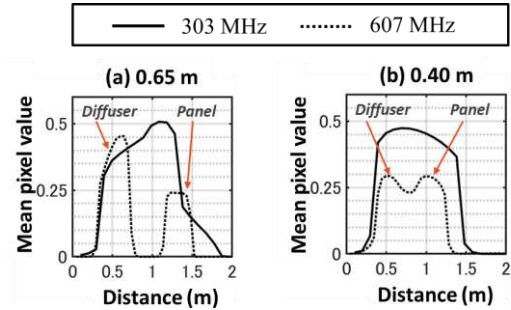
REFERENCES

[1] A. R. Ximenes, *et al.,* "A 256 × 256 45/65nm 3D-Stacked SPAD-Based Direct TOF Image Sensor for LiDAR Applications with Optical Polar Modulation for up to 18.6 dB Interference Suppression", ISSCC 2018, pp. 96–98.

[2] Z. Zhao, *et al.*, "A Novel Imaging Method for Two-Tap Pulsed-Based Indirect Time-of-Flight Sensor," IEEE Sensors Journal, Vol. 23, No. 7, pp. 7017-7030, 2023.

[3] S. Kawahito, *et al.*, "CMOS Lock-In Pixel Image Sensors with Lateral Electric Field Control for Time Resolved Imaging," Proc. 2013 IISW, Vol. 361, 10-6.

[4] K. Kagawa, *et al.*, "A dual mode 303 Megaframes per second charge domain time compressive computational CMOS image sensor," Sensors 2022, Vol. 22.5, 1953.

[5] R. Baranuik, "Compressive sensing [lecture notes]," IEEE Signal Processing Magasize, Vol. 24, No. 4, pp. 118-121, 2007.

[6] C. Li, *et al.*, "An efficient augmented Lagrangian method with applications to total variation minimization," Computational Optimization and Applications, Vol. 56, 2013.
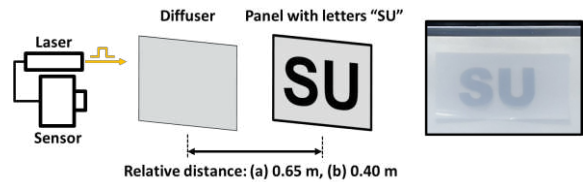
a) Experiment setup for multi-path ToF imaging

b) Reconstructed light waveform at relative distances of 0.65 m and 0.40 m (Average of ROI 5×5 pixel values)

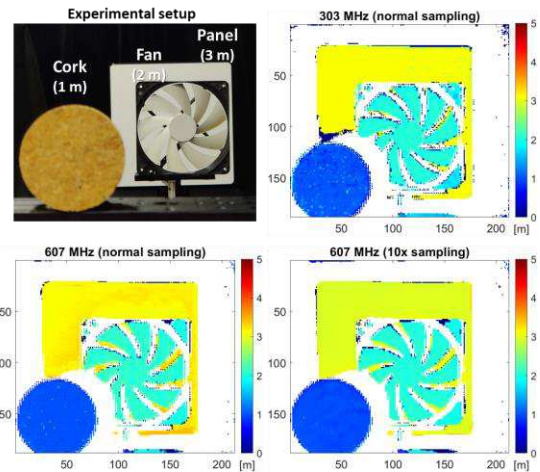Fig. 7. MPI interference separation at 303 MHz and 607 MHz

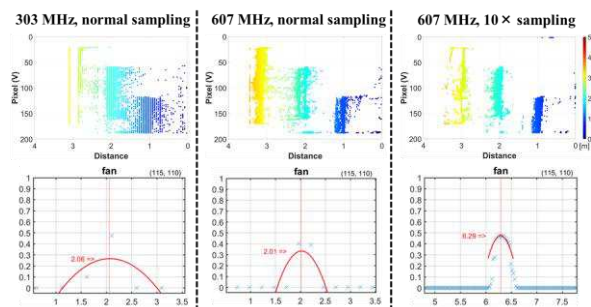Fig. 8. ToF imaging experiment setup and depth maps

Fig. 9. Side view of depth maps and an example of peaks of reconstruction signal

Table 2. Mean and standard deviation of objects' depth (Average of 3×3 ROI pixels of 20 measurements)

	Real depth	303 MHz Normal sampling		607 MHz Normal sampling		607 MHz 10× sampling	
		Mean [m]	Std [cm]	Mean [m]	Std [cm]	Mean [m]	Std [cm]
Cork	1.00	1.04	2.53	1.01	1.17	1.00	0.94
Fan	2.00	1.66	5.59	1.98	1.14	1.99	0.72
Panel	3.00	3.03	3.67	3.08	5.60	3.00	0.57

Histogram-less direct time-of-flight imaging based on a machine learning processor on FPGA

Tommaso Milanese*, Jiuxuan Zhao*, Brent Hearn†, Edoardo Charbon*

*AQUA laboratory, École Polytechnique Fédérale de Lausanne, Neuchâtel

{tommaso.milanese, jiuxuan.zhao, edoardo.charbon}@epfl.ch

†Imaging division, STMicroelectronics, Edinburgh, U.K.

{brent.hearn}@st.com

Abstract—The investigation of a novel architecture for direct time-of-flight (TOF) SPAD based imaging systems is presented. In the proposed architecture, a pulsed laser source illuminates a scene and the reflected light is captured by a SPAD, which detects photons and converts them to a digital pulse. Like in time-correlated single-photon counting (TCSPC), for each detected photon a timestamp is generated, however, unlike TCSPC, it is fed to an machine-learning processor (MLP) that was trained to recognize SPAD responses in direct TOF. The MLP generates the distance to the target directly, taking into account potential non-idealities in timestamp generation and processing. Finally, the proposed architecture was demonstrated in practical scenes and its performance reported using standard LiDAR characterization methods.

I. INTRODUCTION

SPADs are the sensor of choice in many direct TOF and LiDAR systems, thanks to their compactness and picosecond timing resolution, which enables millimetric precision. In TCSPC, timestamps are generated whenever a photon is detected and organized in a histogram. A histogram approaches the true response of the SPAD upon for a very large – ideally infinite – number of detected photons. In practice, an estimate of the TOF is extracted from the histogram after a finite time and thus its precision is also limited. In addition, due to dark noise and background illumination, a typical histogram contains large data, much of it not useful for computing TOF. Hence, the memory allocation for a histogram is generally overestimated and thus inefficient [1], [2]. Indeed, the memory scales exponentially with respect to full scale range (FSR) and hardware timing precision and linearly with the number of depth-dots, leading to a large, possibly on-chip memory [3], [4]. To address this issue, partial histograms have been devised [5]. This approach however, in its simplest embodiment, may prevent the detection of multiple targets at separate depths. To address this shortcoming, more complex partial histograms are needed, along with complex tracking algorithms. An alternative to direct TOF, is the use of indirect TOF and frequency modulated continuous wave (FMCW) techniques, however, these techniques perform averaging at various levels of sophistication, which in effect prevents multiple depth detection as well. In this paper, we propose to use machine learning to process all photon timestamps directly, as soon as they are generated by a time-to-digital converter (TDC) driven by the SPAD image sensor. The objective is the elimination of the histograms needed in a direct TOF configuration, as shown in Fig.1. This approach provides the same advantages
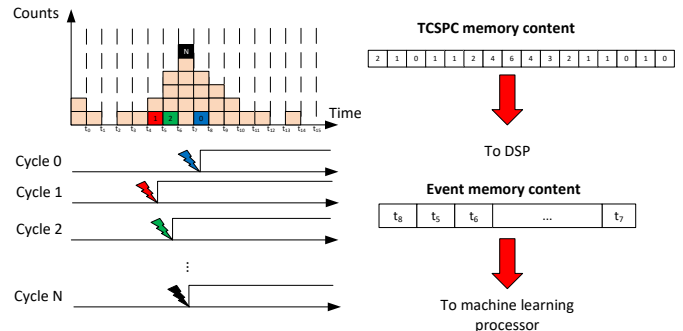

Figure 1. Standard TCSPC and proposed event-processing.

of direct TOF with full histograms, but with much lower requirements in terms of memory and processing power. At the same time, non-idealities associated with the generation and processing of timestamps are intrinsically accounted for. Moreover, the machine-learning processor (MLP) can also be reconfigured to other tasks at a higher level, such as shape and object recognition. The novel machine learning processor was optimized for the long short-term memory (LSTM) execution. We show how to implement the LSTM efficiently in a commercial FPGA, so as to integrate it in the SPAD image sensor in the near future. In the current prototype, photon-detection is performed by a CMOS SPAD, whose raw signals are routed to the FPGA implementing an array of on-demand TDCs, which are then passed on to the LSTM accelerator, as shown in the remainder of the paper.

II. SYSTEM ARCHITECTURE

The general architecture is depicted in Fig.2. The SPAD signal is timestamped by the TDC and the timing data is saved into a $512 \times 32b$ event memory. The host controls the system by means of a state machine, sending a start signal through the Opal Kelly C++ interface and waiting the end of the acquisition and data processing. The LSTM accelerator starts the processing when the event memory is filled, otherwise it stays in sleep mode, reducing the processing power consumption. When the depth retrieval ends the system sends a signal to the host and is ready to get a new depth-dot.
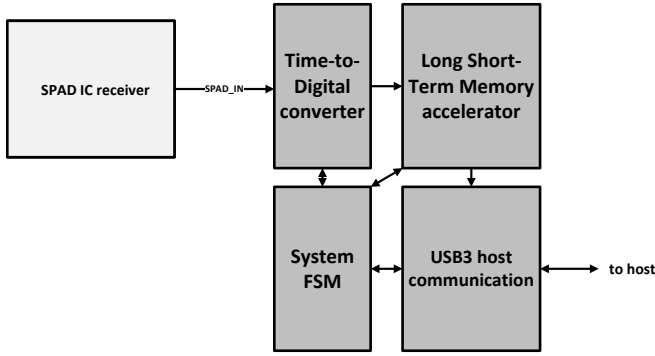
Figure 2. System architecture.

A. Time-to-Digital converter

The TDC is based on a tapped delay-line (TDL) latched at $400\,\text{MHz}$. A chain of Carry4 modules is instantiated in adjacent slices, following the place&route (PnR) tool provided by Xilinx for the ripple carry additions. The thermometer output of the TDL is sampled twice, first by a flip-flop (FF) in the same logic slice as the Carry4 and then by another FF placed by the PnR tool, so as to decrease the probability of metastability in the thermometer code.
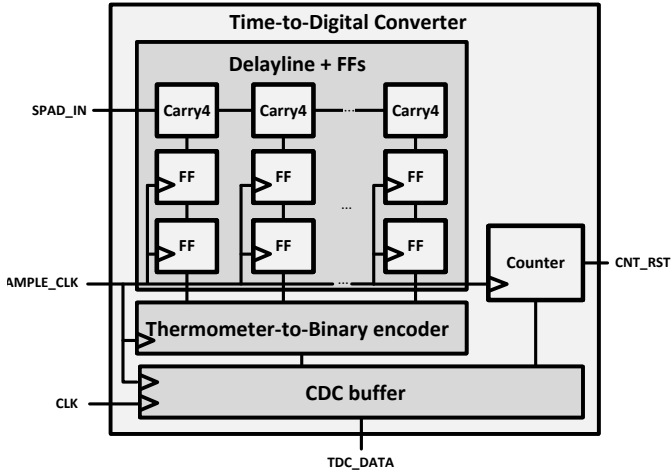

Figure 3. TDC block diagram. Two synchronized clock domains are used operating at 100 and $400\,\text{MHz}$.

A pipelined thermometer-to-binary encoder (T2B) converts the thermometer to binary code at $400\,\text{MHz}$ and passes the data to a clock domain crossing synchronizer, reducing the data rate from the TDL sampling clock to the $100\,\text{MHz}$ system clock. At the end of the signal flow the data is written in the event memory at system clock speed.

B. LSTM accelerator

The LSTM is a recurrent neural network (RNN), a particular type of artificial neural network (ANN) [6]. Since its conception, this processing layer has been used extensively for the processing of time series data, for instance ECG signal classification [7] and speech recognition [8]. The time series

in consideration for SPAD-based D-TOF is the raw timestamp data stream, the same data that is commonly organized in a histogram for peak finding. The governing equations for this network are stated as:

$$f_t = \sigma(\mathbf{W_{xf}}x_t + \mathbf{W_{hf}}h_{t-1} + b_f) \tag{1}$$
$$i_t = \sigma(\mathbf{W_{xi}}x_t + \mathbf{W_{hi}}h_{t-1} + b_i) \tag{2}$$
$$\tilde{c}_t = \tanh(\mathbf{W_{xc}}x_t + \mathbf{W_{hc}}h_{t-1} + b_c) \tag{3}$$
$$o_t = \sigma(\mathbf{W_{xo}}x_t + \mathbf{W_{ho}}h_{t-1} + b_o) \tag{4}$$
$$c_t = f_t \odot c_{t-1} + i_t \odot \tilde{c}_t \tag{5}$$
$$h_t = o_t \odot \tanh(c_t) \tag{6}$$

The bold quantities are matrices and the others are vectors. $\odot$ represents the element-wise multiplication between 2 vectors, $\sigma(\cdot)$ represents the element-wise sigmoid activation for all the vector elements and $\tanh(\cdot)$ is the element-wise hyperbolic tangent activation. After the LSTM layer a final fully connected layer (FCN) is used to transform the final hidden state vector into a regression value, that for this application is a number ranging from 0 to 1 representing the phase of the backscattered light pulse with respect to the laser emitter. This number is then multiplied by the FSR to extract the distance in post-processing. Eqs.1-6 embed matrix-vector multiplications, element-wise additions and multiplications and element-wise non-linear activations, all completely parallelizable operations. The design of this accelerator is based on a row-stationary data flow for the matrix-vector multiplication: each processing elements (PEs) compute one row of the multiplication, acting all in parallel. For the LSTM algorithm execution the resources needed are multipliers, adders, and non-linear activation LUTs, which form the basis of the PE design, Fig.4; muxes are added before the three operators to be able to perform element-wise vector operations. Referring to Eqs.1-6 f_t, i_t, $\tilde{c}_t$, o_t and c_t are stored in the activation registers, h_t in the hidden state memory, all $\mathbf{W}$ and b in the weight memory and x_t in the separate event-memory (not shown in the figure). By reprogramming the PEs scalar and vector calculations can be performed in parallel, lowering the processing time per timestamp. The memories in the design have been laid out in such a way that each PE has its own memory space, both in the weight memory and in the activation registers needed for the LSTM execution. A program counter dictates the state machine execution and, exploiting a masking operation, it controls the address of the weight memory. The weight memory for this design is a 42×128 bit block RAM Xilinx IP, but in the future it will be substituted by a single port SRAM for solid state implementations. The activation registers are implemented with logic, and they are subdivided in two sub-banks per PE in order to decrease the algorithm execution time. The total memory allocated for these registers is 80×28 bits, sub-divided in 10×28 bits for each of the eight PEs.

III. SYSTEM TRAINING AND QUANTIZATION

The LSTM based network has been trained off-line after the creation of a dataset from a MATLAB simulator [9]. The
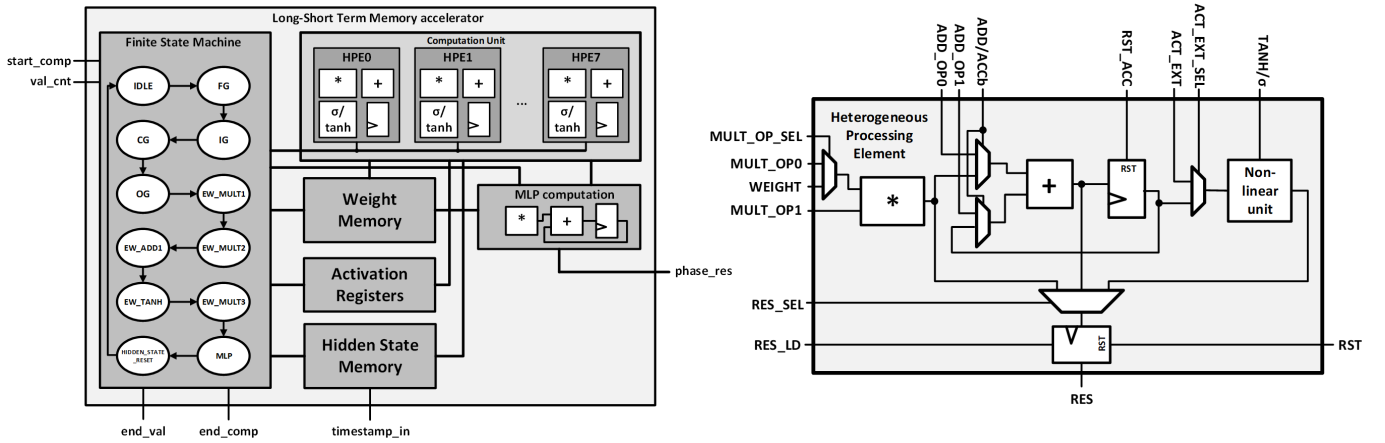
Figure 4. (Left) LSTM accelerator block diagram. (Right) Processing element RTL schematic.

LiDAR signal return is modeled as a Gaussian distribution :

$$P(t;d) = R(d) \exp \left\{ \frac{(t - \frac{2d}{c})^2}{\sigma} \right\}, \qquad (7)$$

where $R(d)$ represents a constant embedding the physical scene/SPAD parameters, d is the target distance, σ is the parameter related to system jitter and c is the speed of light. A background light of 1klux was chosen and a reference white target with 97% reflectivity has been used. The background is assumed to be a uniformly distributed noise source over the laser period with Poisson statistics, where the λ parameter corresponds to the background lux intensity. A gamut of distances ranging from 0 to 15 m with a step of 60 μm was selected and histograms of these distances were constructed and sampled, to obtain 10k time-series per each distance point, containing 512 timestamp values. Fig.5 shows an example of the training data, for the corresponding distance of 8.89 m. The network was implemented in PyTorch and then trained in Google Colab using mean squared error (MSE) loss function:

$$\text{MSE}(y_{p,i}, y_t) = \frac{1}{N_s} \sum_{i=1}^{N_s} (y_{p,i} - y_t)^2, \qquad (8)$$

with N_s being the number of timestamps processed by the network (512 in our case), $y_{p,i}$ the predicted value of the network for every time step and y_t the target ground truth. Note that the ground truth does not change through time, since the whole time series belongs to a single distance distribution, which we want to predict. Adam optimizer with a 0.001 learning rate was used, setting a total of 50 epochs to allow training convergence and input time series have been divided in batches of 64. After training, the network weights were quantized in fixed point arithmetic using simple truncation, in a Q6.10 format. After quantization the weights have been loaded into the weight memory as a memory initialization file for practical reasons, even if a simple DMA has been implemented to allow weight memory reconfigurability on-
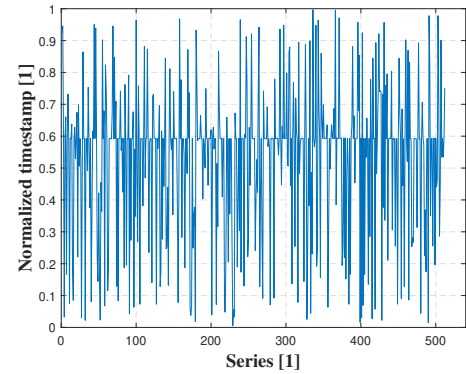
the-fly for solid state implementations.

Figure 5. Example of a simulated time series used for the network training. Time values have been scaled with respect to the maximum range. In this example the distance label is 8.89 m (scaled to 0.593).

IV. RANGING RESULTS

The system-level ranging performance of a single point SPAD sensor has been characterized. The scratchpad buffer processed by the accelerator was loaded with simulated data coming from different probability distributions for different distances, ranging from 0.05 m to 15 m with a step of 60 μm. To simulate a real-life scenario, a scanning confocal setup was built similar to [10] and a statistics of 10,000 points have been acquired for 12 different distances. The field-of-view (FoV) was covered by a generic target consisting of a white paper. The ground truth was acquired with a commercial rangefinder placed below the scanning mirrors, while parallax and offset errors were removed in post-processing. Results for the ranging measurements are shown in Fig.6.

V. 3D IMAGING RESULTS

Using the same optical setup, a 3D image was acquired. The timestamps generated by the TDCs were fed to the MLP and a standard histogram-based center-of-mass (CoM) algorithm. Since the system is event-based, no integration time is set,
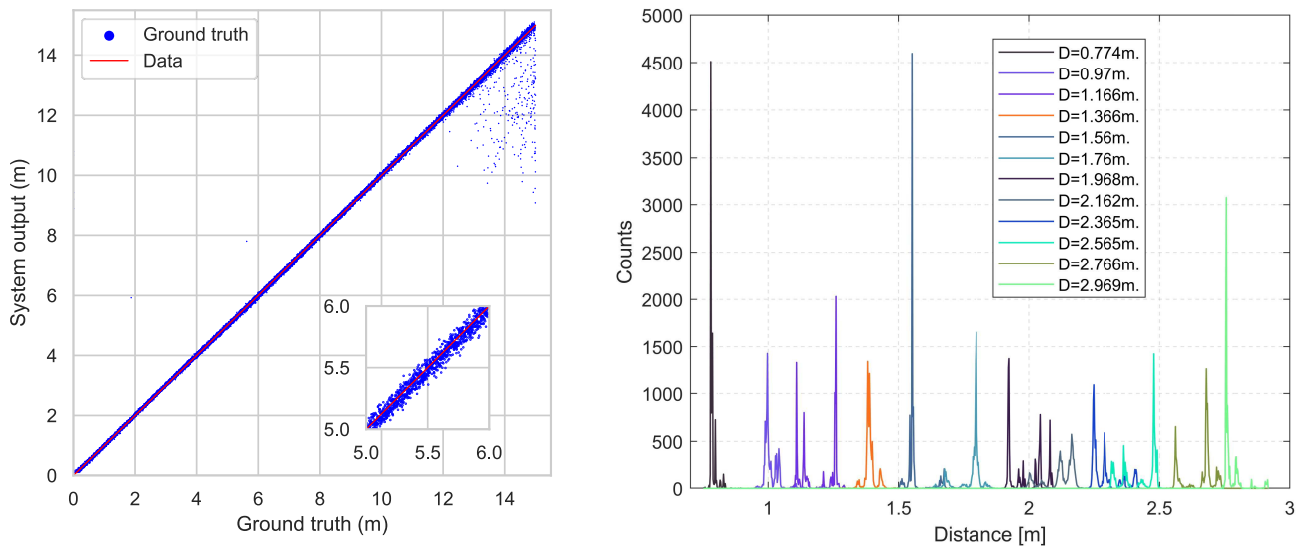
Figure 6. (Left) Simulated data. Input time series step size is 60μm. (Right) Ranging measurements. For each distance 10,0000 points are acquired and organized in a histogram, to characterize the ranging distribution of the implemented LiDAR architecture.

meaning that the acquisition proceeds when the single point has been acquired and processed. To achieve a fair comparison, for each point, we used the same scratchpad buffer where timestamps had been stored for both MLP and CoM. As expected, CoM showed less variability, while the proposed architecture can clearly distinguish the different objects in the scene. The results are shown in Fig.7.
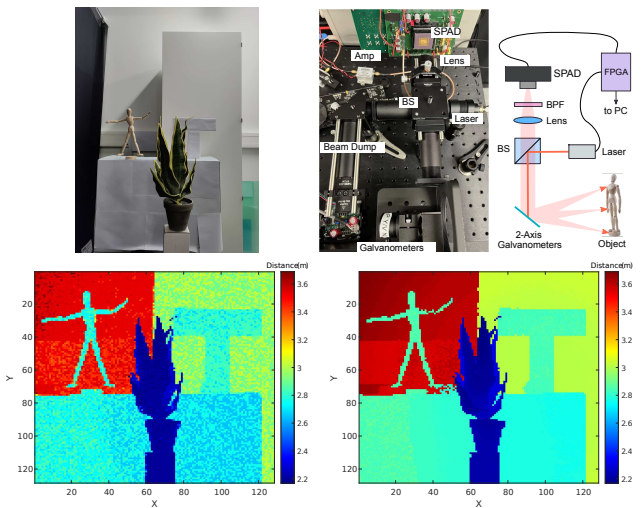

Figure 7. 3D imaging results. Top left: RGB intensity image of the target. Top right: Optical setup. Bottom left: LSTM accelerator. Bottom right: CoM.

VI. CONCLUSIONS

We presented a new histogram-less Direct Time-of-Flight architecture based on timing event processing through a machine learning processor. The processor enables individual photon timestamp processing and it is optimized for LSTM algorithm execution, with the possibility of repurposing it for

other high-speed event-based applications requiring machine learning.

REFERENCES

[1] A. R. Ximenes, P. Padmanabhan, M.-J. Lee, Y. Yamashita, D.-N. Yaung, and E. Charbon, "A 256× 256 45/65nm 3d-stacked spad-based direct tof image sensor for lidar applications with optical polar modulation for up to 18.6 db interference suppression," in *2018 IEEE International Solid-State Circuits Conference-(ISSCC)*. IEEE, 2018, pp. 96–98.

[2] P. Padmanabhan, C. Zhang, M. Cazzaniga, B. Efe, A. R. Ximenes, M.-J. Lee, and E. Charbon, "7.4 a 256× 128 3d-stacked (45nm) spad flash lidar with 7-level coincidence detection and progressive gating for 100m range and 10klux background light," in *2021 IEEE International Solid-State Circuits Conference (ISSCC)*, vol. 64. IEEE, 2021, pp. 111–113.

[3] I. Gyongy, N. A. Dutton, and R. K. Henderson, "Direct time-of-flight single-photon imaging," *IEEE Transactions on Electron Devices*, vol. 69, no. 6, pp. 2794–2805, 2021.

[4] G. Chen, C. Wiede, and R. Kokozinski, "Data processing approaches on spad-based d-tof lidar systems: A review," *IEEE Sensors Journal*, vol. 21, no. 5, pp. 5656–5667, 2020.

[5] C. Zhang, S. Lindner, I. M. Antolović, J. M. Pavia, M. Wolf, and E. Charbon, "A 30-frames/s, 252 × 144 spad flash lidar with 1728 dual-clock 48.8-ps tdcs, and pixel-wise integrated histogramming," *IEEE Journal of Solid-State Circuits*, vol. 54, no. 4, pp. 1137–1151, 2018.

[6] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

[7] S. A. Mirsalari, N. Nazari, S. A. Ansarmohammadi, S. Sinaei, M. E. Salehi, and M. Daneshtalab, "Elc-ecg: Efficient lstm cell for ecg classification based on quantized architecture," in *2021 IEEE International Symposium on Circuits and Systems (ISCAS)*. IEEE, 2021, pp. 1–5.

[8] D. Kadetotad, S. Yin, V. Berisha, C. Chakrabarti, and J.-s. Seo, "An 8.93 tops/w lstm recurrent neural network accelerator featuring hierarchical coarse-grain sparsity for on-device speech recognition," *IEEE Journal of Solid-State Circuits*, vol. 55, no. 7, pp. 1877–1887, 2020.

[9] A. Aßmann, B. Stewart, and A. M. Wallace, "Deep learning for lidar waveforms with multiple returns," in *2020 28th European Signal Processing Conference (EUSIPCO)*. IEEE, 2021, pp. 1571–1575.

[10] J. Zhao, T. Milanese, F. Gramuglia, P. Keshavarzian, S. S. Tan, M. Tng, L. Lim, V. Dhulla, E. Quek, M.-J. Lee *et al.*, "On analog silicon photomultipliers in standard 55-nm bcd technology for lidar applications," *IEEE Journal of Selected Topics in Quantum Electronics*, vol. 28, no. 5, pp. 1–10, 2022.

A 9-shared 3x3 Nonacell Image Sensor with 0.64μm unit pixels for Read Noise and Low-illuminance SNR enhancement

Wonchul Choi, Munhwan Kim, Junoh Kim, Junho Seok, Younguk Song, Dukseo Park, Hyeonseop Yoo, Minyoung Jung, Jieun Lee, Juha Kim, Heegeun Jeong, Kyungho Lee, Eunsang Cho, Howoo Park, Bumsuk Kim, Kyungmin Koh, Sangil Jung, Jungchak Ahn, and Joonseo Yim

Samsung Electronics Co., Ltd., Yongin-city, Gyeonggi-do, Korea, E-mail: wc83.choi@samsung.com

Abstract

A 0.64μm pitch 108 mega pixels CMOS image sensor has been demonstrated and the advanced nonacell structure is used to maximize low-light performance. In this work, a new 9 charge-sum (9S) method is employed, and operation method and parameters for determining signal to noise ratio (SNR) are compared with the conventional 3 charge-sum 3 voltage-average (3S3A). The image characteristics have been examined as the binning method, the enhancement of the device performance due to the generated signal and conversion gain effect have been quantitatively analyzed. Compared to 3S3A, read noise of 9S is reduced by 45%, and improved SNR value is confirmed at low-light operating condition.

Keywords: CMOS image sensor, Nonacell, binning method, charge-sum, read noise, SNR.

Introduction

In recent years, CMOS image sensors (CIS) have attracted great attentions as a mobile application since it is considered as one of the important feature in selecting mobile devices. For this reason, CIS performance improvement is continuously required, and various efforts are underway to improve quality [1]. The demand for high resolution is increasing for instance, however, the pixel size should be shrank due to the limitation of optical format of the lens module [2]. The sub-micron pixel was mass-produced for the first in 2018 [3], and recently, 0.64μm pixel is the smallest one under mass production [4]. However, as the pixel size shrinks down, characteristics such as full well capacity (FWC), sensitivity, and signal-to-noise ratio (SNR) are inevitably degraded. As a way to solve this problem, a pixel summation method such as 2x2, 3x3, and etc. has been proposed and is widely used.

3x3 nona binning is firstly introduced with 0.8μm pixel [5], and it can switch full and binning resolution of 108Mp and 12Mp, respectively. By merging 9 photo-diodes (PDs), especially in low illuminance condition, it improves sensitivity. However, the previous 3x3 nonacell suffered from relatively poor image quality in low illumination condition due to high read noise. In this work, the noise characteristics according to the 3x3 binning method are analyzed in the nonacell structure, and the low-light SNR improvement with new 9 charge sum method is proposed and demonstrated in 0.64μm pixel.

Pixel architecture

In the previous work [5], three 1x3 shared pixel units are merged to generate a 3x3 pixel with the same color. In nona-binning mode, we used 3 charge-sum 3 voltage-average mode (denoted by 3S3A), which indicates that signal electrons in shared 3 PDs are added in the floating diffusion (FD), and the three vertical outputs are averaged at the voltage domain. In this work, on the other hand, 9 charge-sum (denoted by 9S) for
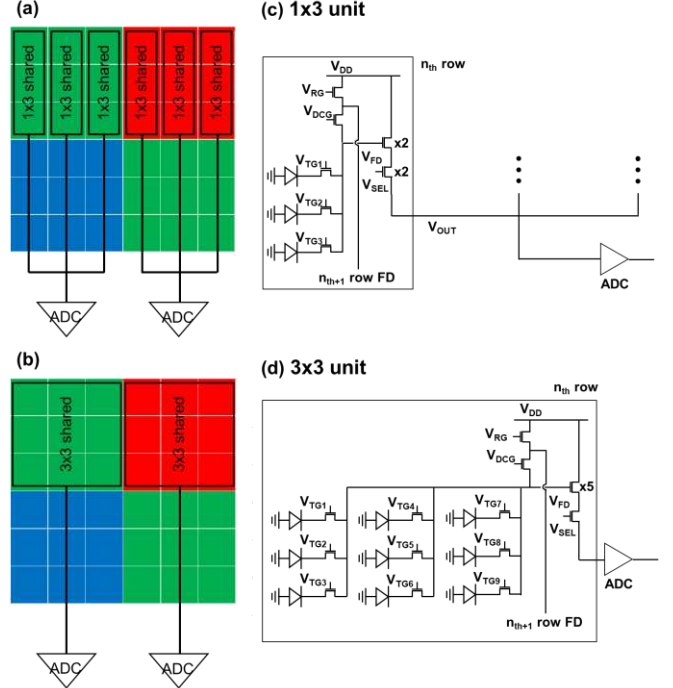

Fig. 1 Nonacell structures with (a) 1x3, (b) 3x3 shared pixels are compared, and the pixel schematic is shown in (c), (d) respectively.

nona-binning is adopted, and the main difference of the 9S between 3S3A is the number of 9 PDs commonly shared to the FD node. The physical structures are compared in Fig. 1(a) and (b).

Since 9 pixels share the same FD, the FWC in binning mode that affects the SNR of high-illumination condition is determined by the FD dynamic range, not the PD's FWC. Therefore, it is important to increase FD capacitance so that it can take as many electrons as possible. In order to make FD capacitance larger, a technique of simultaneously using the capacitance of the adjacent FD node is controlled by reset (RG) and dual conversion gain (DCG) transistors.

TABLE I
3-sum 3-average vs 9-sum mode

	Unit	Full	3S3A	9S
	lsb/e-	$C.G$	$\frac{1}{3} \cdot C.G$	$C.G$
Signal	lsb	S	$3S$	$9S$
Shot noise	lsb	$\sqrt{S}$	$\sqrt{3S}/\sqrt{3}$	$\sqrt{9S}$
SF noise	lsb	N_{SF}	$N_{SF}/\sqrt{3}$	N_{SF}
ADC noise	lsb	N_{ADC}	N_{ADC}	N_{ADC}
S/N ratio	-	$\dfrac{S}{\sqrt{S + N_{SF}^2 + N_{ADC}^2}}$	$\dfrac{9S}{\sqrt{9S + 3N_{SF}^2 + 9N_{ADC}^2}}$	$\dfrac{9S}{\sqrt{9S + N_{SF}^2 + N_{ADC}^2}}$

(a) Calculated SNR

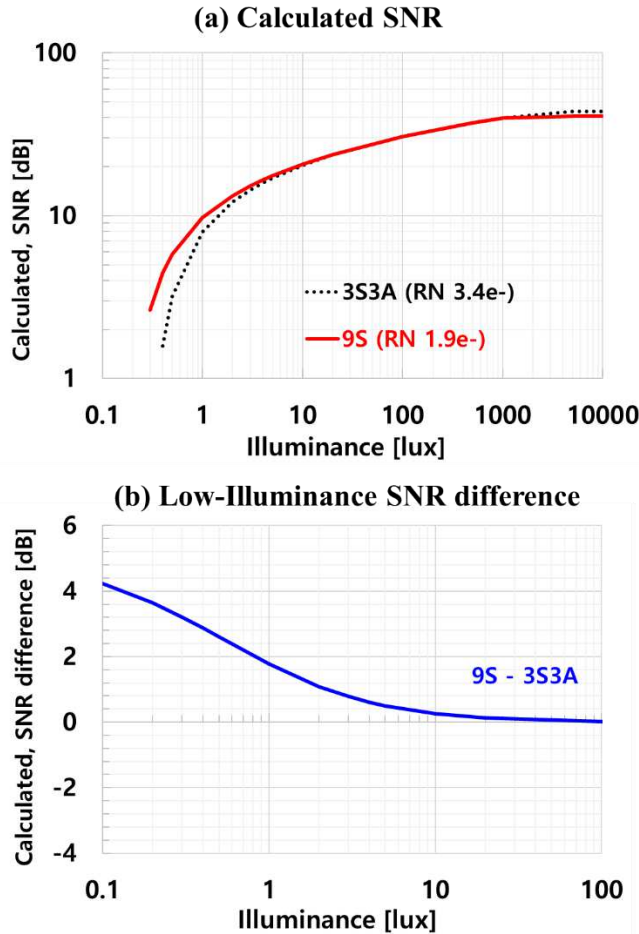## (b) Low-Illuminance SNR difference

Fig. 2 Calculated SNR difference between 9S and conventional 3S3A as a function of illuminance and the same pixel size is assumed.

In the mode where FD capacitance is largely used, the FD node is connected to adjacent 9-shared FD node and total 18-pixels are connected with a single node. Using this method, FD capacitance is increased more than 4 times compared with 9-shared FD capacitance, and the circuit schematic is shown in Fig. 1(c) and (d).

The characteristics of the 9S and conventional 3S3A mode are simply compared in Table I. The conversion gain decreases to 1/3 due to a result of the average, even though the physical FD capacitance is the same. As a consequence, the input referred noise (unit of electron, to be specific) of 3S3A is 3 times larger than that of 9S. In order to compare the SNR between 9S and 3S3A, the signal and noise are analyzed by dividing them in the unit of lsb. Photon shot noise of 9S is relatively large due to its signal, and the dark noise of 3S3A is less than 9S mode since 3S3A averages noise. The signal level takes over the total noise in low-illumination condition ($N \gg S$), and as a consequence, SNR of 9S is superior to that of 3S3A. Fig. 2 shows the differences of SNR between 9S and conventional 3S3A as a function of illuminance. Notice that, for illuminance <10 lux, SNR of 9S is higher than that of 3S3A as we discussed.
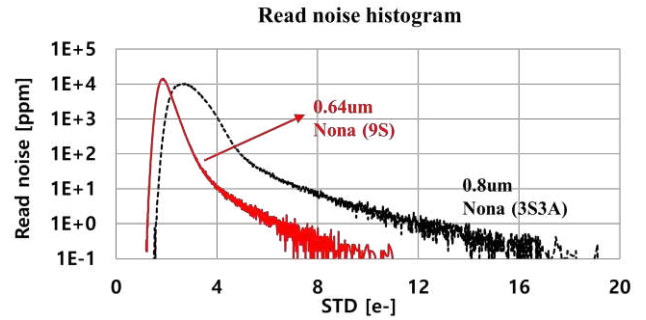
Result and discussion

Fig. 3 The read noise histogram, and 0.64μm for 9S and 0.8μm for 3S3A are compared.

The 9-shared pixel structure has advantages in the placement of transistors, such as source follower (SF), RG, DCG, SEL transistors, in the pixel. As the number of shared pixels increased, there is an additional space for transistors, which is used to apply multi-finger SF. By using multi-finger SF, the channel area of SF is increased, and the width increases as much as it is used in parallel to maximize trans-conductance. As consequence, random telegraph noise of 0.64μm pixel is similar to that of 0.8μm one, even though pixel pitch is reduced down to 0.64μm. Furthermore, it is confirmed that the read noise is reduced from 3.4 to 1.9e-, and the histogram is shown in Fig. 3. Note that read noise in electron unit is increased due to the effect of reducing conversion gain affected by averaging, and it can be represented by $N/\sqrt{3}$ and $\sqrt{3}N$ for lsb and electron units, respectively, where N denotes full mode noise.

The measured characteristics of 9S mode are compared with those of conventional 3S3A as shown in Table II, and we can clearly notice that 9S mode is superior to conventional 3S3A one. SNR characteristic, especially in low illuminance (1lux), is improved due to the read noise. FWC, dark current, and white spot which are key factors in CIS performance remained at the same level as previously published studies [4].

TABLE II
Comparison of pixel characteristics.

	Unit	Tetra	Nona	
		0.64μm	0.8μm	0.64μm
Linear FWC	e-	6,000	6,000	6,000
Dark current	e-/s	1.1	1.5	1.1
Read Noise	e-	1.4	3.4	1.9
RTS	ppm	4	2	0
White spot	ppm	15	10	10
1lux SNR	dB/μm²	4.3	3.3	4.1

White spot : # of pixels≥160lsb @1-frame, gain x16, 200msec, Tj=60℃

RTS : # of pixels≥30lsb @difference between 2-frame, gain x8, 0msec, Tj=60℃

Read noise : 20-frame, gain x16, 33msec, Ta=25℃

Conclusion

In conclusion, we have demonstrated the read noise and SNR characteristics of 9S CMOS image sensor with 0.64μm

unit pixels. A new binning scheme has been introduced to improve SNR at low illuminance, and the 9S and conventional 3S3A are compared in terms of conversion gain, noise, and SNR. The proposed 9 charge sum binning method can dramatically improve read noise, and a 3x3 nonacell structure that can provide image quality equivalent to big pixels at low illumination condition and high resolution at high illumination condition has been completed.

References

[1] I. S. Joe, *2021 Symposium on VLSI Technology, 2021, pp. 1-2* [2] S. Choi, *2017 Symposium on VLSI Technology, 2017, pp. T104-T105* [3] Y. Kim, *2018 IEEE International Solid - State Circuits Conference - (ISSCC), 2018, pp. 84-86* [4] J. E. Park. *2021 IEEE International Solid- State Circuits Conference (ISSCC), 2021, pp. 122-124* [5] Y. Oh, *2020 IEEE International Electron Devices Meeting (IEDM), 2020, pp. 16.2.1-16.2.4*

Light-Emission Crosstalk Model and Dynamic Correction Algorithm for Large-Scale SPAD Image Sensors

A. Abdelghafar, K. Morimoto, H. Sekine, H. Tsuchiya, M. Shinohara, Y. Ota,
J. Iwata, Y. Matsuno, K. Sakurai, and T. Ichikawa

Canon Inc., Kanagawa, Japan,
TEL: +81-3-3758-2111
E-mail: abdelghafar.aymantarek@mail.canon

Abstract—We present a new dynamic correction algorithm to suppress the impact of light-emission (LE) crosstalk in single-photon avalanche diode (SPAD) image sensors. The proposed algorithm demonstrated a capability of correcting both LE crosstalk and nonlinearity in a SPAD sensor, minimizing the effect of hot clusters and color shift that could potentially appear over the array. The results demonstrated a significant improvement in dark signal non-uniformity and color reproducibility. The proposed algorithm allows large-scale SPAD sensors to be implemented in various fields of imaging applications such as surveillance, biomedical and automotive.

I. INTRODUCTION

Single-photon avalanche diode (SPAD) image sensor provides a fully-digital 2D imaging solution, and is a promising candidate for a next generation of photon-counting image sensors for low-light and high dynamic range (HDR) imaging applications. A key challenge in high-definition SPAD cameras is the suppression of light-emission (LE) crosstalk that originates from electron-hole recombination around avalanche multiplication region. Pixel miniaturization leads to enhanced LE crosstalk, and it is thereby critical to mitigate the impact on image quality towards further scaling of the array. In recent decades, many studies have been conducted on the basic principle of LE crosstalk in SPAD arrays [1,2,3]. It has been observed that the spectrum of light emission contains a wide range of wavelengths from visible to near infrared. Unlike typical optical crosstalk and charge crosstalk discussed in CMOS image sensors, LE crosstalk observed in SPAD sensors can generate spurious counts not only to adjacent pixels but to distant pixels. Due to this unique property, there are several issues induced specific to SPAD image sensors. Fig. 1 shows a schematic representation of SPAD pixels with possible LE crosstalk contributions to adjacent pixels. LE crosstalk could cause enhanced hot clusters, increased apparent signal counts or dark counts, color shift, and reduced modulation transfer function (MTF), leading to image quality degradation. Nonlinear photoresponse of SPAD pixels gives additional complexity when estimating a magnitude of LE crosstalk over the array.

To mitigate the impact of LE crosstalk, several approaches have been proposed, including optimization and improvement of process, device, circuit and operation. An important countermeasure is to reduce the amount of LE crosstalk by introducing buried metal full trench isolation between pixels, to prevent emitted light to propagate from one pixel to another [4]. Another approach is the implementation of hot pixel elimination function into pixel circuits [5]. Alternative approach is to adaptively control a frequency of avalanche event through a modified recharging operation [6]. Yet, it is challenging for any of those approaches to thoroughly eliminate the effect of LE crosstalk. To minimize the image quality degradation, an appropriate correction method must be considered. To this date, a systematic method to correct the effect of LE crosstalk in the post-processing has not been presented.

In this paper, we propose a versatile correction algorithm for large-scale SPAD image sensors that suppresses the impact of LE crosstalk. A prototype megapixel SPAD array is used to examine the feasibility of this proposed algorithm.

II. PRINCIPLES & METHODS

A common approach to quantify the level of LE crosstalk is to introduce a crosstalk matrix; a spatial map of crosstalk probability. To obtain the crosstalk matrix for SPAD array, averaged dark frame is captured to extract addresses of isolated hot pixels. For each hot pixel, relative signal levels of the surrounding pixels are mapped around the extracted hot pixel. Based on each signal distribution created from the extracted hot pixels, a spatial map of crosstalk probability can be derived. Crosstalk count at the surrounding pixels is proportional to the signal count at the core hot pixel. This indicates that LE crosstalk contribution is predictable based on the core hot pixel level and the predefined crosstalk map.

Fig. 2(a) illustrates the isolated hot pixels using a synthetic dark image. Due to high reverse bias operation, SPAD image sensors tend to suffer from higher hot pixel population compared to CMOS image sensors. Fig. 2(b) shows an example of LE crosstalk matrix K, representing the crosstalk percentage from central to adjacent pixels. Fig. 2(c) shows the effect of LE crosstalk which can be represented by a convolution of Fig. 2(a) and (b). Each isolated hot pixel behaves as light emitting source to form a hot cluster. Fig. 2(d) shows the result of conventional hot pixel correction (HPC) based on a selective 1D median filtering. LE crosstalk-induced hot clusters spread over multiple pixels, and hence the conventional approach cannot fully eliminate the defective pixels.

To apply a precise correction to LE crosstalk, active recharging-based operation must be employed instead of passive recharging-based operation, where non-monotonic photoresponse precludes a unique

estimation of crosstalk contribution. As one of the examples of active recharging, clocked recharging is conducted by providing a periodic recharging clock fed to a gate of recharge transistors [6]. In contrast to the conventional passive recharging, the clocked recharging-based sensor restricts the maximum photon counts per frame allowing to reduce power consumption, which makes it a viable option for megapixel resolution sensor suited for imaging applications.

Fig. 3(a) and (b) represents a cross-section, a pixel circuit, and a timing diagram of clocked recharging-based SPAD sensor; a photo-electron triggers avalanche multiplication which induces LE crosstalk to adjacent pixels. Fig. 3(c) depicts a relation between the number of detected avalanche events (N_{ct}) and the number of incident photons (N_{ph}) for active and passive recharging-based operations. In passive recharging, (N_{ct}) decreases at high-light condition. In clocked recharging, one recharging event allows only one avalanche event at maximum, leading to a nonlinear monotonic photoresponse.

Fig. 4 shows a signal conversion flow for the clocked recharging-based SPAD pixels in the presence of LE crosstalk. The number of LE events is equivalent to N_{ct}, and contribution of the secondary photons is fed back to N_{ph} in the adjacent pixels. This feedback process could induce higher-order crosstalk events, and complicates analytical formulation of LE crosstalk contribution. However, the correction algorithm for LE crosstalk can be simplified when the aforementioned crosstalk matrix K is introduced.

Fig. 5 shows an image processing flow of the proposed LE crosstalk correction algorithm. First, a raw image is converted to "Image A" through nonlinear correction function F^{-1}. In parallel, the same raw image is convolved with matrix K' to create "Image B", where matrix K' represents a contribution of the LE crosstalk, obtained by replacing the central element of matrix K with 0. "Image B" is subtracted from "Image A" to create "Image C", which shows the result of both nonlinear correction and LE crosstalk correction applied. Finally, "Image C" is converted to "Image D" through conventional HPC. The proposed correction algorithm for LE crosstalk is represented by a following equation:

$$I_t = F^{-1}(I_{raw}) - I_{raw} * K'$$

where I_{raw} is the initial raw image, and I_t is the reconstructed image. Benefit of this algorithm is that it can simultaneously correct nonlinearity and LE crosstalk-induced artifacts, such as hot cluster and color shift, irrelevant to light intensity distribution.

To evaluate the feasibility of the proposed algorithm, a variant of 3.2Mpixel 3D-stacked back-illuminated SPAD sensor is used, which has relatively higher crosstalk probability than those presented in a previous report [7].

III. RESULTS

Fig. 6 shows the results of conventional HPC and the proposed correction applied to a 100-frame-averaged dark image captured by the 3.2Mpixel SPAD image sensor. In contrast to the conventional HPC, the proposed algorithm shows a significant improvement in the dark signal non-uniformity. Fig. 7 illustrates the diagram of measured hot pixel population over the array in each image presented in Fig. 6. The hot pixel population (>50cps at 25□) of the proposed algorithm has been reduced by more than 6 times with respect to the conventional HPC method.

Fig. 8 shows the images obtained using a SPAD sensor with on-chip Bayer color filter in mid-light (top images) and high-light (bottom images) conditions. Fig. 8(a) shows raw images, representing their relative difference in light intensity. Fig. 8(b) and (c) show the result of nonlinearity correction applied, and the result of LE crosstalk correction additionally applied, respectively. Fig. 8(b) shows a color shift towards magenta, where a degree of shift is dependent on the light level. Fig. 8(c), in contrast, shows robust color reproduction for different light levels. This result indicates that the proposed correction algorithm can properly compensate LE crosstalk-induced nonlinear color shift.

IV. CONCLUSION

In this paper, a dynamic correction algorithm to reduce the effect of LE crosstalk is presented. The algorithm demonstrated a significant suppression of LE crosstalk-induced hot cluster and color shift, in the presence of nonlinear photoresponse. The proposed algorithm allows the implementation of multi-megapixel SPAD image sensors into various fields of imaging and sensing applications, including electronic industry, surveillance, biomedical and automotive.

V. REFERENCES

[1] I. Rech *et al.*, "Optical crosstalk in single photon avalanche diode arrays: A new complete model," *Opt. Express*, 16 (12) 8381-8394, 2008.

[2] R. Younger *et al.*, "Crosstalk analysis of integrated Geiger-mode avalanche photodiode focal plane arrays," *Defense + Commercial Sensing*, 2009.

[3] S. Jahromi and J. Kostamovaara, "Timing and probability of crosstalk in a dense CMOS SPAD array in pulsed TOF applications," *Opt. Express*, 26 (16), 20622-20632, 2018.

[4] K. Ito *et al.*, "A Back Illuminated 10μm SPAD Pixel Array Comprising Full Trench Isolation and Cu-Cu Bonding with Over 14% PDE at 940nm," *IEEE Int. Electron Devices Meeting (IEDM)*, 16.6.1-16.6.4, 2020.

[5] Y. Maruyama *et al.*, "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*, 49 (1), 179-189, 2014.

[6] Y. Ota *et al.*, "A 0.37W 143dB-Dynamic-Range 1Mpixel Backside-Illuminated Charge-Focusing SPAD Image Sensor with Pixel-Wise Exposure Control and Adaptive Clocked Recharging," *IEEE Int. Solid- State Circuits Conference (ISSCC)*, 94-96, 2022.

[7] K. Morimoto *et al.*, "3.2 Megapixel 3D-Stacked Charge Focusing SPAD for Low-Light Imaging and Depth Sensing," *IEEE Int. Electron Devices Meeting (IEDM)*, 20.2.1-20.2.4, 2021.
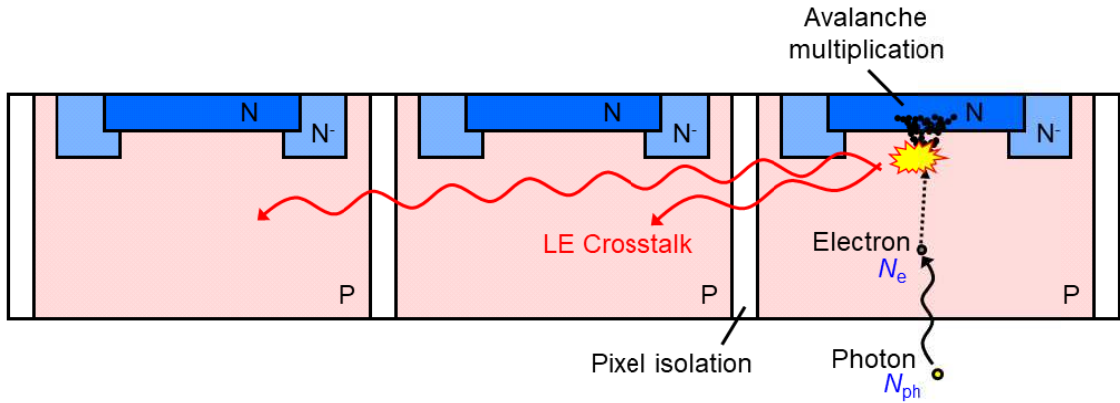
Fig. 1. Schematic representation of SPAD pixels and LE crosstalk contributions from one pixel to its adjacent pixels.
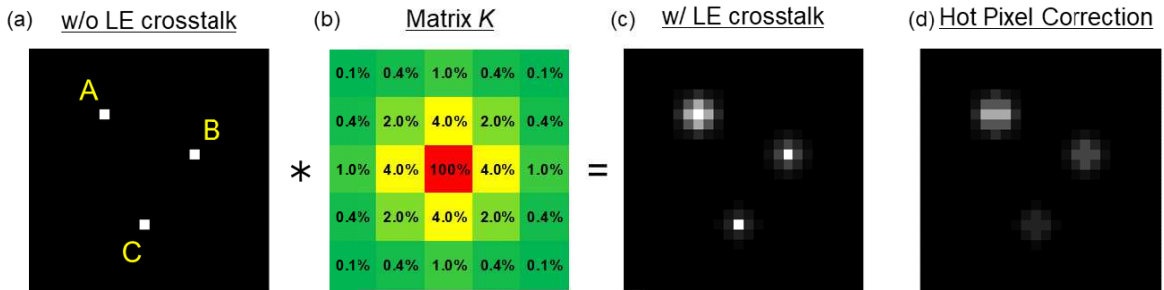

Fig. 2. Schematic views describing impact of LE crosstalk on image quality. (a) Synthetic image representing isolated hot pixels. (b) Conceptual example of 5×5 LE crosstalk matrix showing LE crosstalk percentage from central to adjacent pixels. (c) Synthetic image simulating the effect of LE crosstalk applied on (a). (d) Result of conventional hot pixel correction on (c) based on selective 1D median filtering.
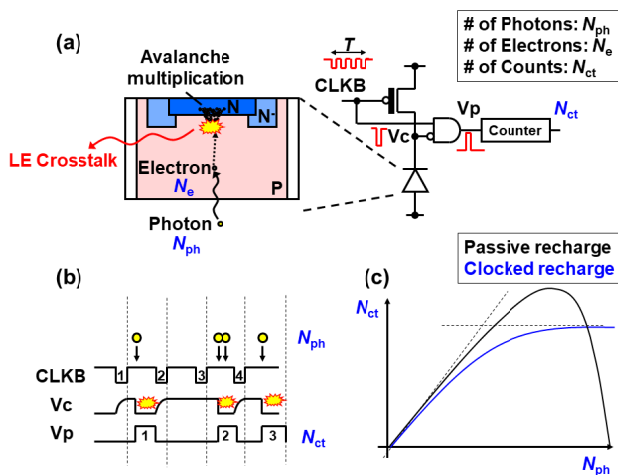
Fig. 3. Conceptual views of clocked recharging-based operation. (a) Cross-section and pixel circuit of SPAD pixel. (b) Timing diagram. (c) Schematic plot of relation between N_{ct} and N_{ph} for clocked and passive operations.
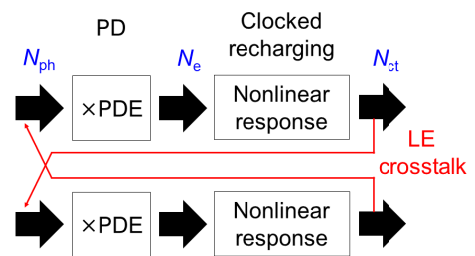
Fig. 4. Signal conversion flow for SPAD pixels in the presence of nonlinearity and LE crosstalk.
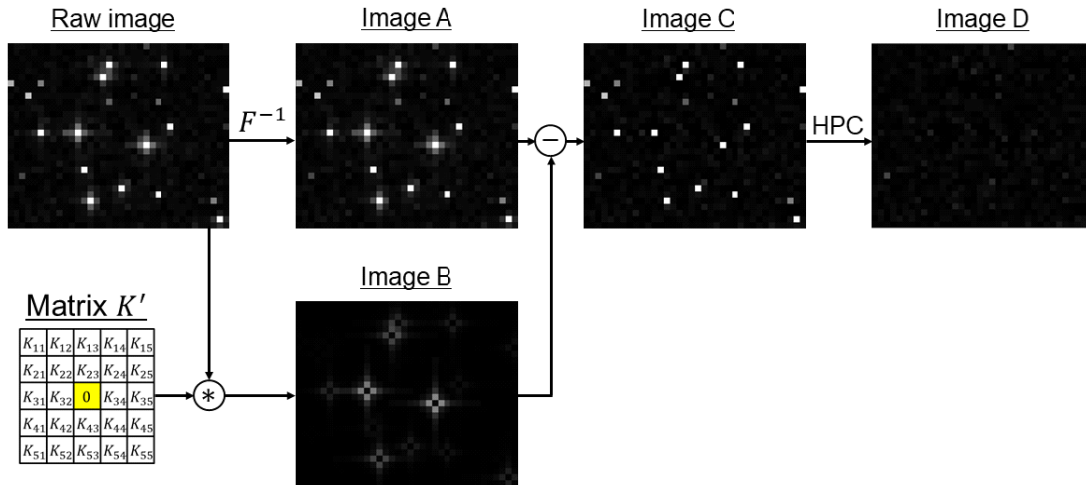
Fig. 5. Image processing flow of the proposed LE crosstalk correction algorithm.
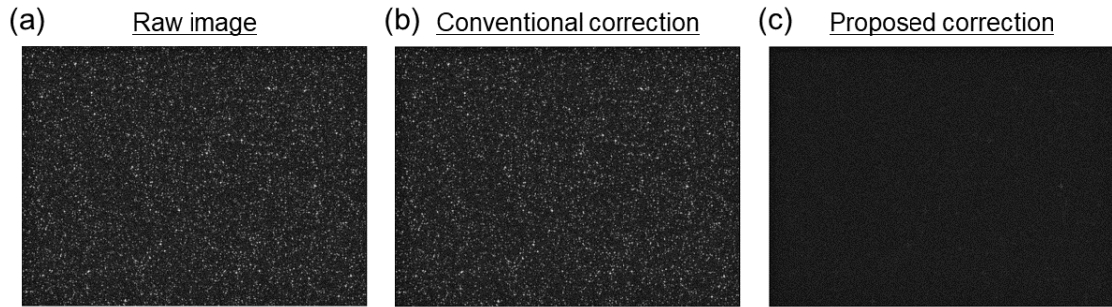
Fig. 6. Results of LE crosstalk correction. (a) Averaged full-resolution dark image captured by 3.2Mpixel SPAD image sensor. (b) Result of the conventional hot pixel correction. (c) Result of the proposed LE crosstalk correction.
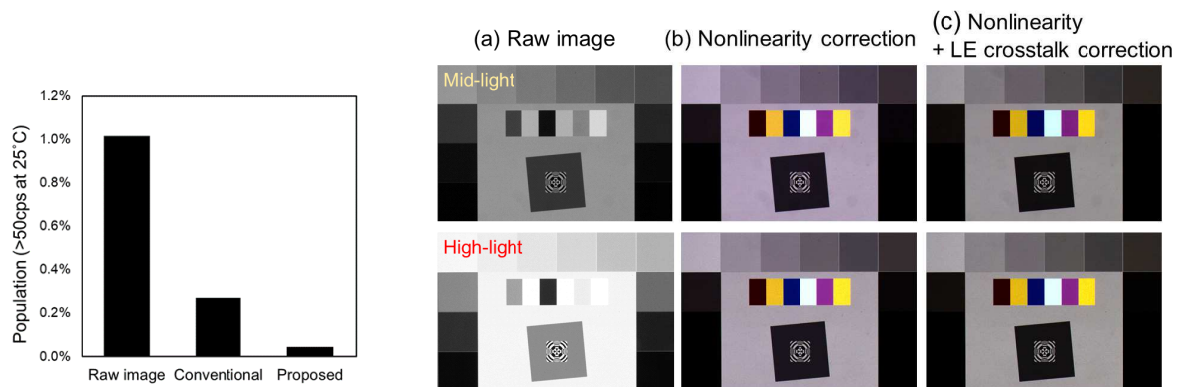

Fig. 7. Diagram of measured hot pixel population for raw image, conventional correction and the proposed correction.

Fig. 8. Results of nonlinearity correction and LE crosstalk correction in color image under mid-light (top images), and high-light (bottom images) conditions. (a) Raw images captured by 3.2Mpixel SPAD image sensor. (b) Results of nonlinearity correction. (c) Results of the LE crosstalk correction simultaneously applied. Digital gain is applied to compare the color tone under equivalent contrast for (b) and (c).

Optimal biasing and physical limits of DVS event noise

Rui Graca, Brian McReynolds, Tobi Delbruck

Sensors Group, Inst. of Neuroinformatics, UZH-ETH Zurich, Zurich, Switzerland

rpgraca,bmac,tobi@ini.uzh.ch, https://sensors.ini.uzh.ch

Abstract—Under dim lighting conditions, the output of Dynamic Vision Sensor (DVS) event cameras is strongly affected by noise. Photon and electron shot-noise cause a high rate of non-informative events that reduce Signal to Noise ratio. DVS noise performance depends not only on the scene illumination, but also on the user-controllable biasing of the camera. In this paper, we explore the physical limits of DVS noise, showing that the DVS photoreceptor is limited to a theoretical minimum of 2x photon shot noise, and we discuss how biasing the DVS with high photoreceptor bias and adequate source-follower bias approaches optimal noise performance. We support our conclusions with pixel-level measurements of a DAVIS346 and analysis of a theoretical pixel model.

I. Introduction

The Dynamic Vision Sensor (DVS) [1]–[4] is a neuromorphic event-based vision sensor, which consists of an array of asynchronously operating pixels as the one in Fig. 1 [4]. Each pixel independently encodes instantaneous changes in its input light into an asynchronous steam of ON and OFF events. More specifically, a pixel outputs an ON event when the relative Temporal Contrast (TC) [1] of light intensity at its input increases by a user defined ON threshold since the last event, or an OFF event when the relative TC increases by a user defined OFF threshold since the last event. When its input is static, a DVS pixel ideally outputs no event. More extensive description of the DVS pixel operation can be found in [1], [5], [6].

Characteristics of the DVS such as sparse data encoding and low latency make it a good candidate for scientific applications such as space situational awareness and wide-field voltage and calcium imaging. The adequacy of the DVS for some applications is potentially limited by a too high rate of parasitic Background Activity (BA). BA consists of events that do not encode changes in the input. These events are undesirable because they decrease the Signal-to-Noise Ratio (SNR) and increase data volume [5], [8]. The BA of the DVS pixel strongly depends on both light intensity and camera biasing [3], [6], [9]–[11]. It is predominantly caused by photon and electron shot noise in dark settings [9], and by leakage in the reset transistor (Fig. 1F) in brighter settings [10].

A good understanding of the phenomena resulting in BA is important for improving camera models that can aid pixel design, optimization of the camera utilization, or learning algorithms [5], [6], [12], [13].

In [9], noise power at the output of the photoreceptor (V_{pr} in Fig. 1) and noise event rate are explored as a function of illumination and photoreceptor bias I_{pr}. There, we observe that both noise power and event rate are lower for lower I_{pr}. This occurs because the bandwidth is lower for lower I_{pr}.

These observations suggest that using a small I_{pr} to limit bandwidth reduces noise, and this assumption has been used as an optimization rule for bias control [13]. In this paper, we go a step further into understanding the optimal conditions and biasing of the DVS pixel, and show that in fact the opposite is generally true – even though strongly reducing I_{pr} leads to a decrease in noise events, noise performance is more optimal for high I_{pr}. We show that the DVS photoreceptor topology is bounded with a theoretical minimum of 2x photon shot noise, and we discuss bias optimization regarding bandwidth and its implications on noise and signal. In this paper, we focus on the biasing of the photoreceptor (Fig. 1**A**) by I_{pr} and the Source-Follower buffer (SF) (Fig. 1**B**) by I_{sf}. A more general discussion about bias optimization is presented in [6], and considerations about threshold and refractory biases are discussed in [11].

II. Optimal Photoreceptor biasing

A. PSD Measurements and modeling

Fig. 2**a** shows the noise PSD measured at V_{pr} of a test pixel isolated from a DAVIS346 array under an on-chip illuminance of $0.1\,\mathrm{lx}$ for two different I_{pr} settings: one high ($3\,\mathrm{nA}$) and one low ($10\,\mathrm{pA}$). The dashed lines in the figure show the PSD predicted by a theoretical physically-realistic model operating under the same conditions. The theoretical model was obtained by circuit analysis considering the sources of shot noise in the photoreceptor and applying the transfer function that relates them to V_{pr}. The parameters for the model were then estimated and fitted based on SPICE simulation and pixel measurements.

Since the theoretical model generally matches both measured and simulated data, we utilize it to further infer about the noise contribution of each noise source to the total output noise. In Fig. 2**b**, we see how the contribution of the photocurrent I_{pd} (depicted by the dotted lines and consisting of photon shot noise at the photodiode and electron shot noise added by M_{fb}) and the contribution of I_{pr} (depicted by the dashed lines, and consisting of noise introduced by M_n and the transistor implementing I_{pr}) add up to the total PSD. Here, we observe that the level of the contribution of I_{pd} is independent of I_{pr}, but its bandwidth may depend on I_{pr} – for a bias of $10\,\mathrm{pA}$, I_{pr} is right at the edge of starting to filter out the I_{pd} contribution. That is, this contribution would be
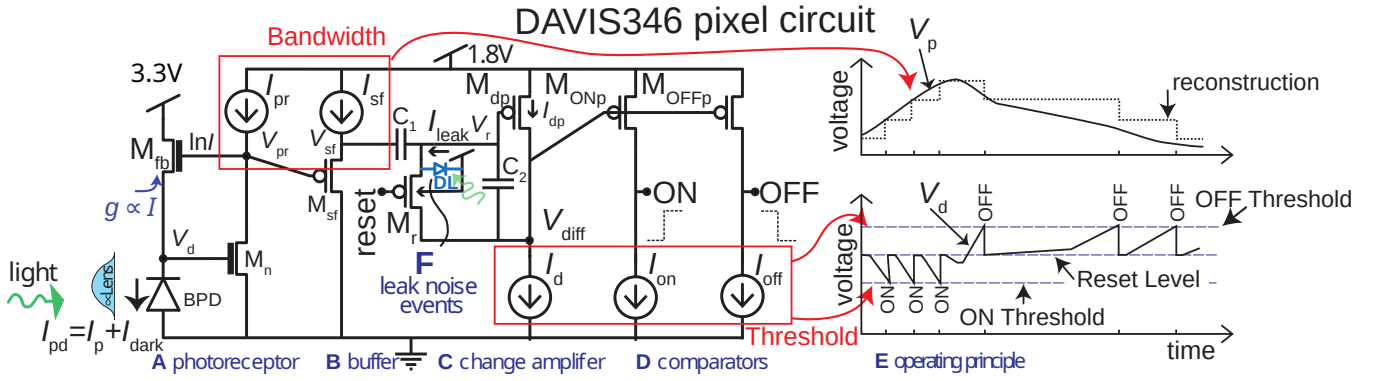
Fig. 1. Typical DVS pixel circuit [7]. The active logarithmic photoreceptor (**A**) is buffered by a source-follower (**B**), which drives a cap-feedback change amplifier (**C**), which is reset on each event by a low-going *reset* pulse. A finite refractory period holds the change amplifier in reset for the refractory period Δ_{refr}. Comparators (**D**) detect ON and OFF events as seen in **E**. Periodic leak events result from junction and parasitic photocurrent I_{leak} in diode DL (**F**).

significantly reduced for lower I_{pr}, and would become constant for higher I_{pr} (as happens for I_{pr} of 3 nA). On the other hand, the contribution of I_{pr} moves to higher frequencies when I_{pr} increases.

Fig. 2**c** shows the square root of the integral of the different components of the PSD in Fig. 2**b**. The final value of the square root of the integral is the RMS voltage noise contribution of its respective source. We see that the contribution of I_{pr} converges to a value independent of I_{pr} - the contribution is only shifted to higher frequencies. The contribution of I_{pd} is lower for lower I_{pr}, which happens due to filtering by I_{pr} [9]. For higher values of I_{pr}, filtering would stop occurring and the contribution of I_{pd} converges to the constant value observed at I_{pr} of 3 nA.

Figs. 2**e** and 2**f** show the modeled PSDs and the square root of their integrals for the contributors at the output of the SF, V_{sf}. The PSDs were obtained by filtering the ones at V_{pr} using a model of the SF estimated by circuit inspection and simulation. Also, the noise contribution of I_{sf} is added in the dashed line. However, its value is much smaller than the contribution of the photoreceptor (the summation of the contributions of I_{pd} and I_{pr}).

Fig. 2**d** shows the modeled signal transfer function from logarithmic changes in light intensity to voltages at V_{pr} and V_{sf}. As described in [9], it can be approximately modeled as a second order system with one pole dependent on I_{pd} and the other dependent on I_{pr}. At I_{pr} of 3 nA, the pole controlled by I_{pd} is clearly dominant, while for I_{pr} of 10 pA the two poles lie very close to each other. The SF add another pole, which for the bias used is close to the dominant of the photoreceptor.

Fig. 3 show the noise rates measured from the same test pixel for varying I_{pr} for two different on-chip illumination levels. We observe that for high I_{pr}, noise rate becomes mostly constant, since all the noise components of I_{pr} are filtered out. For middle I_{pr} values, the noise contributions of I_{pr} lie within the signal bandwidth and are not filtered out, and the noise rates peak. For lower I_{pr} the noise rates decrease because I_{pr} limits the bandwidth.

B. Optimal biasing and optimality analysis

From Fig. 2**c** we can see how strongly biasing I_{pr} results in shifting the noise components added by I_{pr} to higher frequencies outside the bandwidth of interest for signal. This means that we can filter them out using the SF without consequences for signal. In the limit, if we bias I_{pr} so strongly that all its contribution is removed by SF, the output noise consists of only the noise contribution of I_{pd} (which consists itself on equal parts of photon shot noise and M_{fb} noise), and the much smaller noise contribution of I_{sf}. In this case, we are theoretically limited to a minimum of 2x photon shot noise when the contribution of I_{sf} becomes negligible.

The clear advantage of strongly biasing I_{pr} is illustrated in Fig. 2**f**. For I_{pr} of 3 nA, the model predicts a contribution of photon shot noise of 46% (approximating the theoretical limit of 50%), resulting in a noise event rate of 0.02 Hz under nominal threshold and refractory biases [6] versus 12% for I_{pr} of 10 pA, resulting in a noise event rate of 0.66 Hz.

The model predicts an RMS noise contribution equivalent to TC log-e units of 0.006 for I_{sf}. The contributions of I_{pd} and I_{pr} depend on filtering, but for the case where the pole controlled by I_{pd} is dominant and filtering by the SF is not considered, they are respectively 0.04 and 0.06 for most values of I_{pd} and I_{pr}. Although RMS noise alone is not enough to characterize DVS noise, since it does not contain information about the noise frequency [9], these numbers are useful to evaluate design limitations to the event sensitivity (i.e. the minimum event threshold with acceptable noise rates). One important conclusion is that I_{sf} should be adjusted to the minimum acceptable bandwidth for each application and I_{pr} should be adjusted so that all its contributions are filtered out. Given that increasing I_{pr} increases power consumption, I_{pr} should be optimized to trade off power with noise performance. In the limit where the photoreceptor bandwidth is much higher than the SF bandwidth (which happens for very high illuminance, high I_{pr} and nominal or low I_{sf}) the noise, noise introduced by I_{pd} and I_{pr} is filtered out and SF becomes the main noise contributor.
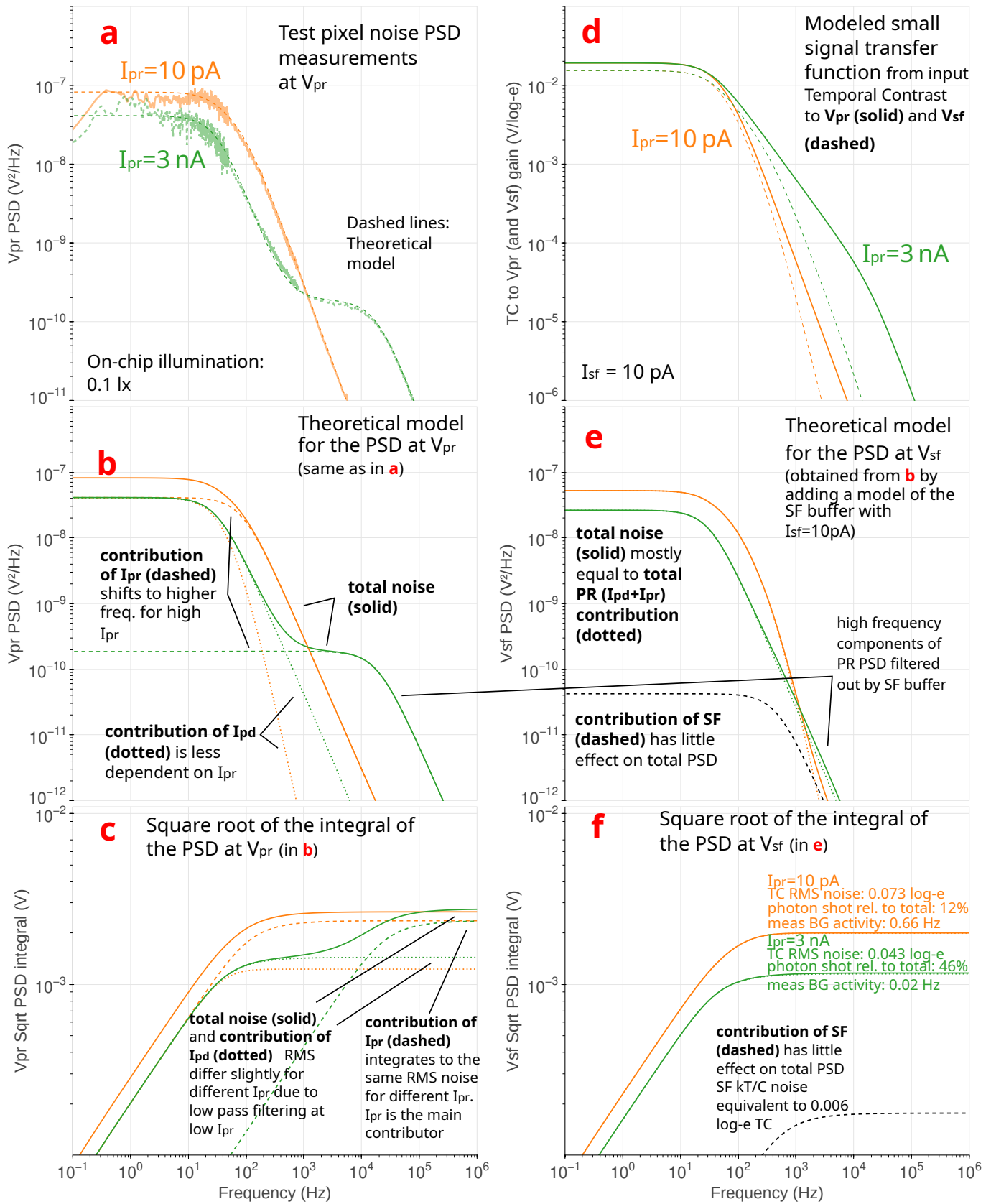
Fig. 2. **(a)** shows noise PSDs measured from a DAVIS346 test pixel for two different I_{pr} biases for an on-chip illuminance of 0.1 lx and the PSDs estimated by a theoretical model for the same conditions. **(b)** shows the estimated contributions of I_{pr} and I_{pd} to the total PSD for the same model in the same conditions, and **(c)** shows the square root of the integral of the curves in **(b)**. The final value of these curves is the respective contribution to the RMS noise voltage at V_{pr}. **(e)** and **(f)** show the same quantities as **(b)** and **(c)**, but relative to V_{sf}. **(d)** shows the estimated signal transfer function from TC (in log-e units) to V_{pr} and V_{sf}.
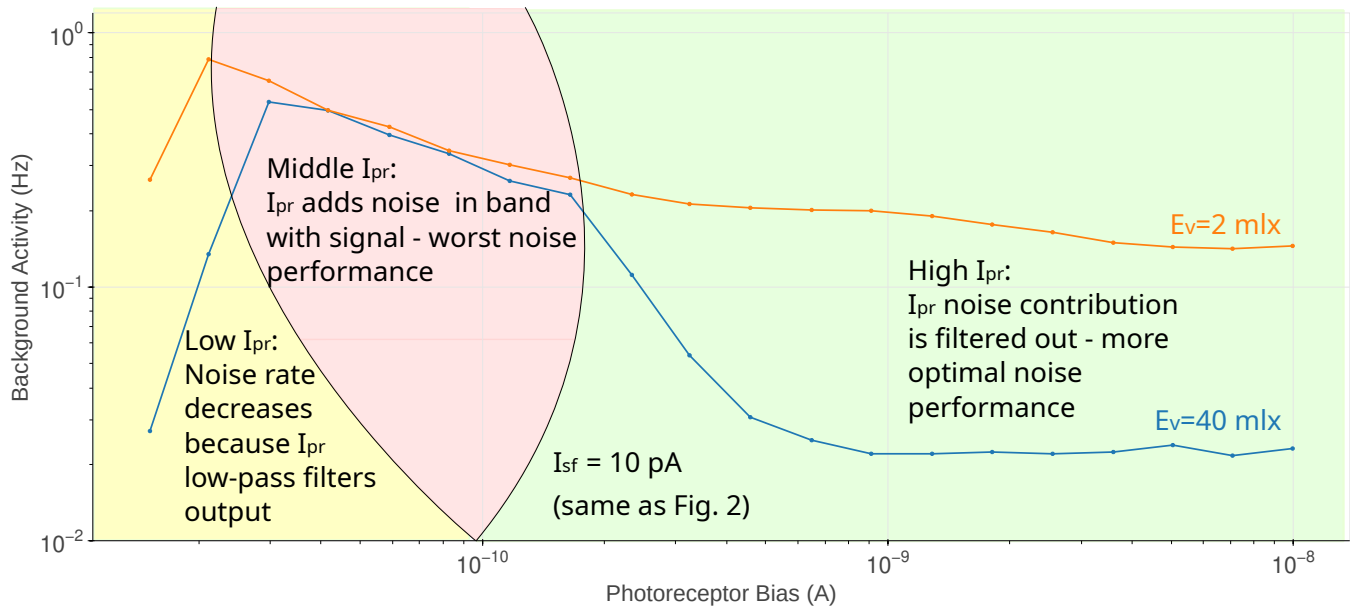
Fig. 3. Background activity measured from a DAVIS346 test pixel under constant on-chip illuminance of $2\,\mathrm{mlx}$ (orange line) and $40\,\mathrm{mlx}$ (blue line) for I_{sf} of $10\,\mathrm{pA}$ (as in Fig. 2) and nominal threshold and refractory bias settings (see [6] for a characterization of these parameters, nominal settings corresponds to tweaks of 0 there).

Filtering with SF and not with I_{pr} is generally a better idea since it introduces significantly less noise, and the noise it introduces is not filtered out in any case. However, in practical DVS implementations operating in very dark settings, very low I_{pr} may result in a lower bandwidth than the minimum achievable by the SF, and minimizing both I_{pr} and I_{sf} may result in less BA.

III. CONCLUSION

The measurements and analysis presented show that the DVS pixel is limited to a minimum of 2x photon shot noise, and that using high I_{pr} and adequate I_{sf} approximates this limit. We also discuss the limits imposed to event sensitivity by each noise contributor.

REFERENCES

[1] P. Lichtsteiner, C. Posch, and T. Delbruck, "A 128×128 120 dB 15 μs latency asynchronous temporal contrast vision sensor," *IEEE Journal of Solid-State Circuits*, vol. 43, no. 2, pp. 566–576, 2008.

[2] Y. Suh, S. Choi, M. Ito, *et al.*, "A 1280×960 dynamic vision sensor with a 4.95-μm pixel pitch and motion artifact minimization," in *2020 IEEE International Symposium on Circuits and Systems (ISCAS)*, 2020, pp. 1–5. DOI: 10.1109/ISCAS45731.2020.9180436.

[3] T. Finateu, A. Niwa, D. Matolin, *et al.*, "5.10 a 1280×720 back-illuminated stacked temporal contrast event-based vision sensor with 4.86µm pixels, 1.066GEPS readout, programmable event-rate controller and compressive data-formatting pipeline," in *2020 IEEE International Solid- State Circuits Conference - (ISSCC)*, 2020, pp. 112–114. DOI: 10.1109/ISSCC19947.2020.9063149.

[4] C. Brandli, R. Berner, M. Yang, S. Liu, and T. Delbruck, "A 240 × 180 130 dB 3 μs latency global shutter spatiotemporal vision sensor," *IEEE Journal of Solid-State Circuits*, vol. 49, no. 10, pp. 2333–2341, 2014. DOI: 10.1109/JSSC.2014.2342715.

[5] Y. Hu, S.-C. Liu, and T. Delbruck, "V2e: From video frames to realistic DVS events," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, Jun. 2021, pp. 1312–1321.

[6] R. Graca, B. McReynolds, and T. Delbruck, "Shining light on the DVS pixel: A tutorial and discussion about biasing and optimization," in *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, Jun. 2023. DOI: 10.48550/arXiv.2304.04706.

[7] G. Taverni, D. Paul Moeys, C. Li, *et al.*, "Front and back illuminated dynamic and active pixel vision sensors comparison," *IEEE Transactions on Circuits and Systems II: Express Briefs*, vol. 65, no. 5, pp. 677–681, 2018.

[8] S. Guo and T. Delbruck, "Low cost and latency event camera background activity denoising," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 45, no. 1, pp. 785–795, 2023. DOI: 10.1109/TPAMI.2022.3152999.

[9] R. Graca and T. Delbruck, "Unraveling the paradox of intensity-dependent DVS pixel noise," in *2021 International Image Sensor Workshop (IISW)*, Sep. 2021. DOI: 10.48550/ARXIV.2109.08640.

[10] Y. Nozaki and T. Delbruck, "Temperature and parasitic photocurrent effects in dynamic vision sensors," *IEEE Transactions on Electron Devices*, vol. 64, no. 8, pp. 3239–3245, Aug. 2017, ISSN: 1557-9646. DOI: 10.1109/TED.2017.2717848.

[11] B. McReynolds, R. Graca, and T. Delbruck, "Exploiting alternating DVS shot noise event pair statistics to reduce background activity," in *2023 International Image Sensor Workshop (IISW)*, May 2023. DOI: 10.48550/arXiv.2304.03494.

[12] B. J. McReynolds, R. P. Graca, and T. Delbruck, "Experimental methods to predict dynamic vision sensor event camera performance," *Optical Engineering*, vol. 61, no. 7, p. 074 103, 2022. DOI: 10.1117/1.OE.61.7.074103.

[13] T. Delbruck, R. Graca, and M. Paluch, "Feedback control of event cameras," in *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2021, pp. 1324–1332. DOI: 10.1109/CVPRW53098.2021.00146.

Metasurface-based planar microlenses for SPAD pixels

Jérôme Vaillant
Univ. Grenoble Alpes
CEA, LETI
38000 Grenoble, France
jerome.vaillant@cea.fr

Lucie Dilhan
STMicroelectronics
Imaging division
Grenoble, France
lucie.dilhan@st.com

Alain Ostrovsky
STMicroelectronics
TR&D
Crolles, France
alain.ostrovsky@st.com

Quentin Abadie
Univ. Grenoble Alpes
CEA, LETI
38000 Grenoble, France
quentin.abadie@cea.fr

Lilian Masarotto
Univ. Grenoble Alpes
CEA, LETI
38000 Grenoble, France
lilian.masarotto@cea.fr

Romain Paquet
Univ. Grenoble Alpes
CEA, LETI
38000 Grenoble, France
romain.paquet@cea.fr

Mickaël Cavelier
Univ. Grenoble Alpes
CEA, LETI
38000 Grenoble, France
mickael.cavelier@cea.fr

Cyril Bellegarde
Univ. Grenoble Alpes
CEA, LETI
38000 Grenoble, France
cyril.bellegarde@cea.fr

Abstract— **In this paper we present two design generations of metasurface-based planar microlenses implemented on Front-Side Illumination SPAD pixels. This kind of microlens is an alternative to conventional reflow microlens. It offers more degrees of freedom in term of design, especially the capability to design off-axis microlens to gather light around the SPAD photodiode. The two generations of microlenses have been fabricated on STMicroelectronics SPAD and characterized. We validated the sensitivity improvement offered by extended metasurface-based microlens. We also confirmed the impact of lithography capability on metasurface performances, highlighting the need have access to advance deep-UV lithography.**

Keywords— metasurface, planar microlens, pixel, SPAD

I. INTRODUCTION

Planar microlenses are well suited for SPAD pixels, which usually work under monochromatic illumination (Time-of-Flight or Fluorescence-Lifetime applications) and exhibit rather large pixel size with lower fill-factor compared to state-of-the-art CMOS pixels. We previously demonstrated Fresnel Zone Plate lenses[1] and proposed metasurface based[2] microlenses[3], [4]. In this paper, we present the design, fabrication and characterization results of two generations of metasurface-based planar microlenses on STMicroelectronics SPAD[5] array.

II. META-ATOM AND PLANAR MICROLENS DESIGN

Our unitary structure, or meta-atom, is a nanoscale pillar of high refractive index material (amorphous silicon) embedded in a low refractive index medium (silicon oxide). The phase shift induced by a pillar is controlled by its geometry which is defined by the pitch, the capping thickness and the pillar's parameters: height and diameter (see Fig. 1).
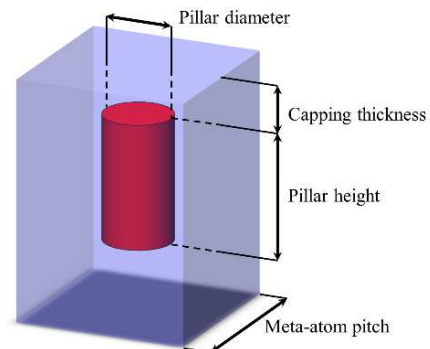

Fig. 1. Geometry of meta-atom defined by the meta-atom pitch, the thickness of silicon oxide capping and the pillar height and diameter

We have also considered the paving strategy. For the first generation of metasurface-based microlens, we considered square paving for meta-atom arrangement. In order to improve the spatial sampling, triangular paving (see Fig. 2) was also implemented in the second generation.
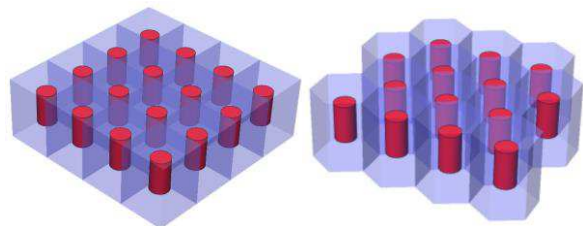

Fig. 2. Paving geometries used to design metasurface: square paving (left) and triangular paving (right)

We defined pillar libraries as sets of pillar with the paving and the first three parameters fixed and the

pillar's diameter varies inside the range achievable by lithography: the minimal diameter is defined by the minimal Critical Dimension (CD) and the maximal diameter is defined as meta-atom pitch minus the minimal space. Among all possible libraries, we select the ones that cover a phase shift from 0 to 2π while offering the best transmission (see Fig. 3).
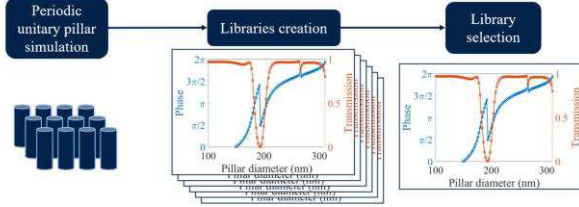

Fig. 3. Library generation and selection workflow

First generation of libraries is based on 500nm and 420nm pitches whereas the second generation is more aggressive with meta-atom pitch of 370nm. In both case, the minimum CD and space considered are 100nm.

III. PLANAR MICROLENS DESIGNS

A. SPAD pixel

For this development, we consider 32×32 SPAD arrays. SPAD are sharing N-well and thus are grouped 4×4 into a $86.4 \times 86.4 \mu m^2$ cell (see Fig. 4). The SPAD itself has a dimension of $10.5 \times 11.5 \mu m^2$. So, we use the capability to design off-axis microlens with metasurface to extend the footprint compared to conventional refractive reflow-based microlens (i.e. $10.5 \times 11.5 \mu m^2$) and thus collect more light.
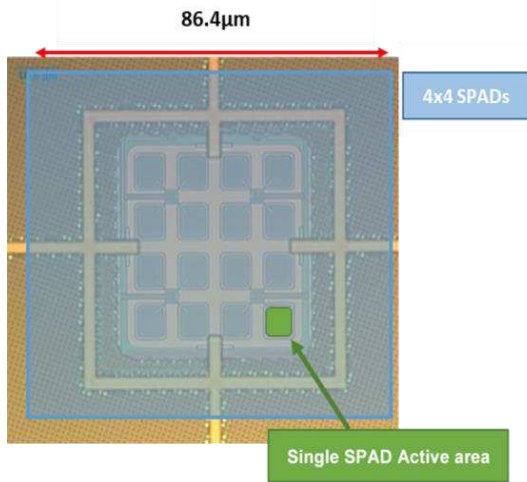
Fig. 4. Layout of 4x4 SPAD cell sharing the same N-well

B. Planar microlens

We divide the 32×32 array of SPAD into 8 areas of 8×16 SPAD. Each area is covered by a given design of microlens (see Fig. 5):

- One area without any microlens, to get the bare SPAD sensitivity as reference

- One area with a microlens having the same footprint as the reflow one ($S_1 = 10.5 \times$

$11.5 \mu m^2$). This will be used for direct comparison with refractive microlens.

- Three area with microlens having intermediate footprint $S_{2.25} = 16.5 \times 16.5 \mu m^2$, which is 2.25 times larger than the reference surface.

- Three area with microlens having the largest possible footprint $S_{3.86} = 21.6 \times 21.6 \mu m^2$, 3.86 times larger than the reference.
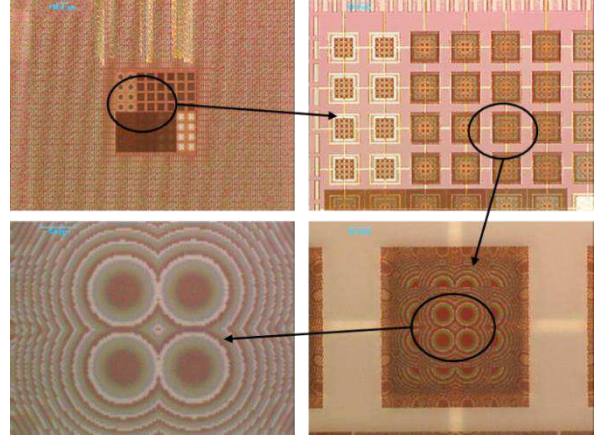

Fig. 5. Views of 32x32 SPAD array coverd by microlenses, with a focus on microlens footprint x2.25

It should be noted that for microlenses with footprint larger than the reference, the design differs according to their position inside the 4x4 group of SPAD (see bottom left picture in Fig. 5). Considering layout symmetry, we distinguish four different pixels: center, top, corner and side (see Fig. 6).
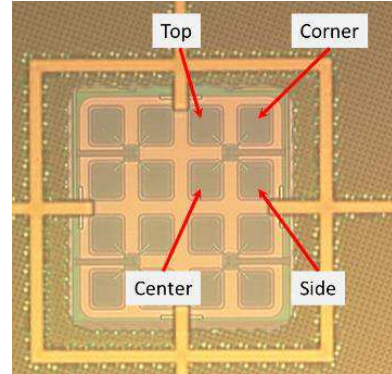

Fig. 6. Naming of SPAD and microlens according to their position in the 4x4 group

Microlens optical axis is centered on the SPAD of interest. When microlens and SPAD have the same footprint the optical axis is centered on it. But when the microlens surface extends outside SPAD footprint, the optical axis is offset (see bottom left image on Fig. 5). The target phase profile of the metasurface corresponds to a perfect lens bending a plane wave into a spherical one:

$$\varphi(x,y) = \frac{2\pi}{\lambda}\sqrt{(x-x_0)^2 + (y-y_0)^2 + f^2} - f \quad (1)$$

Spatial coordinates are denoted x and y, the offset of the optical axis of the microlens x_0 and y_0, the focal length

of the microlens *f*, and the wavelength of interest λ (940nm in our case). To encode this phase profile we consider the classical look-up table method[4].

IV. METASURFACE FABRICATION

The process flow starts on *40nm* CMOS Front-Side Illumination SPAD wafers[5] with the optical pedestal (SiO_2) deposition and planarization. Then, a low-stress layer of amorphous silicon (aSi) is deposited and planarized. The meta-atom are defined by dry deep-UV lithography and etching. Finally, SiO_2 deposition and planarization ensure the pillar's encapsulation and the capping thickness is tuned to minimize the reflection of the metasurface (see Fig. 7).
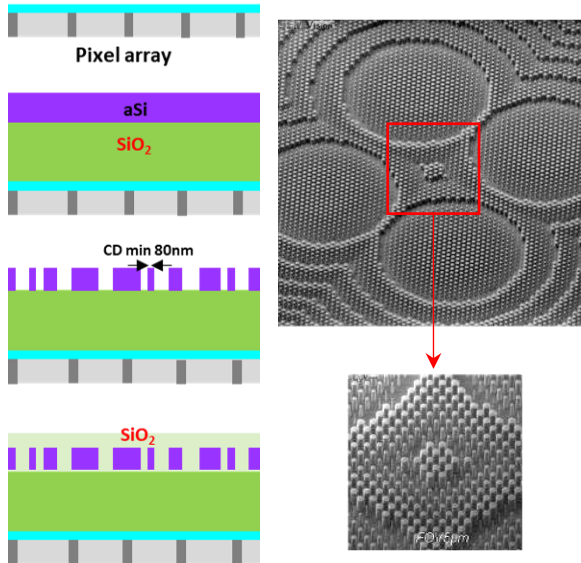
Fig. 7. Schematic of process flow (left) and SEM tilted view of planar metasurface based-microlens (right)

V. ELECTRO-OPTICAL CHARACTERIZATION

A. Experimental setup

The electro-optical characterization of the SPAD arrays are done at wafer-level with dedicated probe station based on Accretech 300mm prober. The light source is a Thorlabs M940L3 LED filtered with a Thorlabs bandpass filter FBH940-10 ($FWHM = 10nm$). Uniform illumination over the array is ensured by a LabSphere integrating sphere. Fixed distance between the sphere output and the wafer emulate an f/10 angular distribution. The light intensity is recorded using calibrated radiometer (UDT 221).

To evaluate microlens performances, we calculate the Photon Detection Efficiency (PDE), i.e. the quantum efficiency:

$$PDE = \frac{LCR - DCR}{\Phi_{940nm} \times a_{SPAD}^2} \quad (2)$$

With Φ_{940nm} the optical flux in $photon.m^{-2}.s^{-1}$ at the surface of the wafer, a_{SPAD}^2 the surface of the SPAD pixel (i.e. $21.6 \times 21.6 \mu m^2$), LCR the Light Count Rate and DCR the Dark Count Rate, which are the frequency of SPAD triggering respectively under illumination and in darkness. As these count rates usually depend of the

excess bias, above the breakdown voltage of the SPAD, we first evaluate the mean breakdown voltage of each circuit. The excess bias is set to 1.5V above the breakdown voltage for LCR and DCR measurement.

B. Experimental results

Both generation of metasurface-based microlens have been characterized as well as SPAD with refractive microlens (process of reference) and bare SPAD without microlens. The Fig. 9 shows the PDE for all of these configurations.

As expected, for bare SPAD and for microlens with unitary surface ($S_1 = 10.5 \times 11.5 \mu m^2$) the PDE are almost the same whatever the SPAD. Indeed, the layout of the 4 kind of SPAD (center, top, corner, side) are almost the same over this surface S_1. In this configuration, the refractive microlens leads to higher sensitivity than metasurface-based microlens as it bends the phase in a continuous way, contrary to metasurfaces that sample the phase profile, spatially and in phase value.

For extended surfaces ($S_{2.25}$ and $S_{3.86}$), we clearly see that the PDE depends on the position (center, top, corner, side): the central ones (dots in Fig. 9) are more sensitive than the corner ones (diamons in Fig. 9). There are two root causes to this effect. Firstly, the layout under a given microlens depends on its position: corner microlens is mainly above metallic interconnections that surround the 4×4 photodiodes, and layout below top and side microlenses are not exactly identical. Secondly, for the phase given by (1), the greater the distance to optical axis, the greater the slope of the phase. So as the meta-atom spatially sample the phase, this leads to possible aliasing. With the more aggressive design rules (smaller pitch and CD and use of triangular paving) the second generation of metasurfaces have higher and less dispersed performances.
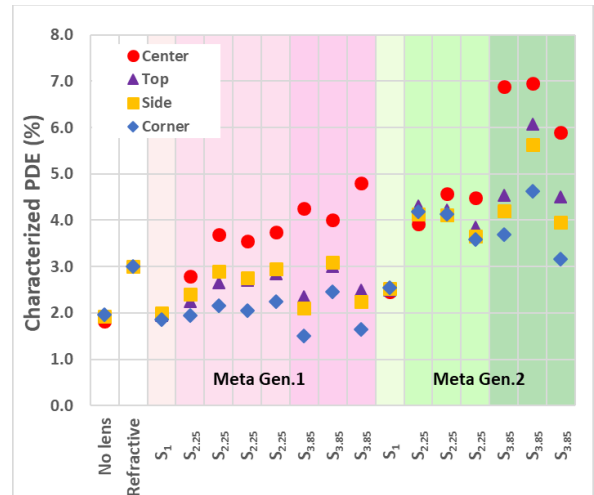

Fig. 8. PDE measurement for for bare SPAD (No lens), with refractive microlens (Refractive) and metasurface-based microlens with footprint of S_1, $S_{2.25}$ and $S_{3.86}$

Compared to reflow microlens, the PDE is improved whatever the metasurface-based microlens. The sensitivity is improved by 30% with small dispersion for our first extended design. For the largest

microlens ($S_{3.86}$) we improve the PDE by $\times 2.3$ for the center microlens, and in any case the PDE is higher than reference (reflow microlens).

VI. CONCLUSION

This work validates the interest and the feasibility of metasurface-based microlens at pixel level. We demonstrate our capability to process deep-subwavelength pillars of amorphous silicon encapsulated in silicon dioxide on top of Front-Side Illumination CMOS wafer to generate microlenses. Measurement on 32x32 SPAD array confirms the interest of such technology. With the capability to design off-axis microlenses, we take advantage of space available around the 4x4 SPAD group to improve the PDE compared to the classical reflow microlenses.

ACKNOWLEDGMENT

REFERENCES

[1] J. Vaillant *et al.*, « SPAD array sensitivity improvement by diffractive microlens », présenté à International Image Sensor Workshop, 2019, p. 4.

[2] P. Genevet, F. Capasso, F. Aieta, M. Khorasaninejad, et R. Devlin, « Recent advances in planar optics: from plasmonic to dielectric metasurfaces », *Optica*, vol. 4, nº 1, p. 139-152, janv. 2017, doi: 10.1364/OPTICA.4.000139.

[3] L. Dilhan, J. Vaillant, A. Ostrovsky, L. Masarotto, C. Pichard, et R. Paquet, « Planar microlenses for near infrared CMOS image sensors », *Electron. Imaging*, vol. 2020, nº 7, p. 144-1-144-7, janv. 2020, doi: 10.2352/ISSN.2470-1173.2020.7.ISS-144.

[4] L. Dilhan *et al.*, « Planar microlenses applied to SPAD pixels », in *IISW proceedings*, 2021.

[5] S. Pellegrini *et al.*, « Industrialised SPAD in 40 nm technology », in *2017 IEEE International Electron Devices Meeting (IEDM)*, déc. 2017, p. 16.5.1-16.5.4. doi: 10.1109/IEDM.2017.8268404.

Flexible Spectrally-Scanning Snapshot Multispectral Imaging On Dual-Tap Coded-Exposure-Pixel CMOS Image Sensors

Roberto Rangel, Navid Sarhangnejad, Mian Wei, Rahul Gulve, Gairik Dutta, Zhengfan Xia, Nikita Gusev, Nikola Katic, Harel Haim, Kiriakos N. Kutulakos, and Roman Genov

Abstract— We present a method of spectrally-scanning snapshot multispectral imaging (MSI) that employs a dual-tap coded-exposure-pixel (CEP) CMOS image sensor. A frame exposure time is divided into N subexposures. During each subexposure, an arbitrarily programmable exposure code is sent to each pixel to control the integration of the photogenerated charge into one of the two taps. We employ the data-memory pixel (DMP) architecture for the CEP, which achieves the smallest pixel size or all CEP sensors. Five unique-wavelength LEDs are sequentially turned on, synchronously with five unique 2x2-pixel code tiles, and submitted to the sensor over five subexposures. The sorted photogenerated charges are read out, and five images at the five wavelengths are subsequently extracted by demultiplexing. The number of wavelengths is flexible and can be easily extended using a larger pixel tile. As a result, spectra for a scene are captured at 5 wavelengths in the visible light and NIR spectrum in a single frame, at 30 frames per second, without using a color filter array.

I. INTRODUCTION

Image classification by modern inference methods such as deep neural networks (DNNs) has surpassed human capabilities in several applications, such as face recognition and skin cancer detection. Spectral sensors, including multispectral imaging (MSI) cameras, yield additional visual information that can further boost the image classification performance in a wide range of applications where spectral information, beyond RGB, is present.

MSI cameras operate by sensing light in a small number of spectral bands, with common applications such as aerial surveillance and crop/food inspection. Two general classes of 2-D MSI cameras exist: (1) spectrally scanning cameras and (2) non-scanning, or snapshot, cameras. The former has slow operation speed that leads to motion artifacts when the incident light is changing rapidly; and the latter suffer from high computational demands and cost. For example, a typical spectrally scanning MSI camera is realized by operating a monochrome sensor in conjunction with several optical filters or illumination sources [1], each with a different wavelength, which are used one per frame, requiring multiple frames in order to acquire one image.

We present a method of spectrally-scanning snapshot MSI that offers the best of both worlds, eliminating the disadvantages of each. It employs an image sensor with a dual-tap coded-exposure pixel (CEP) [2-6] which enables single-shot operation with low computational complexity and cost. The emerging class of CEP image sensors has already been demonstrated to offer superior performance, particularly in the presence of rapidly changing incident light, with a wide range of novel capabilities, such as single-shot compressive sensing [3], and single-shot HDR and 3D imaging [5]. In the presented method, a dual-tap CEP image sensor is utilized to perform single-shot multispectral imaging, with a programmable number of arbitrary wavelengths spanning the visible and near-infrared light spectrum.

High-speed cameras can be employed for spectrally-scanning snapshot MSI, but they have limited output frame rate and suffer from high power, high read noise and high output data rate making them very expensive and thus of limited utility [7]. To avoid some of these drawbacks, general-purpose non-integrated CEP imaging systems that could be suitable for spectrally-scanning snapshot MSI have been developed that employed digital micromirror devices (DMDs) or liquid crystals on silicon (LCoS) to either pass or block light coming to each single-tap pixel of a camera depending on a digital "code" for that pixel [8]. Such systems offer lower readout power, lower read noise, and lower output data rate, as the photogenerated charge is accumulated over multiple coded intra-frame subexposures and is read out only once per frame, as compared to high-speed cameras where a readout takes place for each exposure and thus contributes to higher

power dissipation, read noise and output data rate [5]. However, such non-integrated systems require bulky, expensive and distortive optical components.

CMOS-based multi-tap CEP image sensors [9] offer the additional key advantages of not only smaller form factor and lower cost due to pixel programmability directly in a CMOS Image Sensor (CIS) technology, but also better optical signal fidelity, since no external optical devices or moving parts are needed; and better light efficiency, as the photogenerated charge is sorted among multiple taps instead of being discarded when the one-tap pixel in non-integrated CEP imaging systems is "off".

II. SYSTEM IMPLEMENTATION

The CEP image sensor employs the state-of-the-art dual-tap coded-exposure pixel architecture we refer to as the data-memory pixel (DMP) architecture. Figure 1 (left) highlights the detailed principle of operation of the DMP sensor. During the sensor's image acquisition operation, the exposure period of each frame is divided into N subexposures which are performed before a single readout is made. While photogeneration for the current subexposure takes place, the photogenerated charge from the previous subexposure, temporarily stored on a light-shielded pinned storage diode, referred to as the data memory, is being sorted between two taps row by row, based on each pixel's binary coefficient referred to as the exposure code, or simply the code.
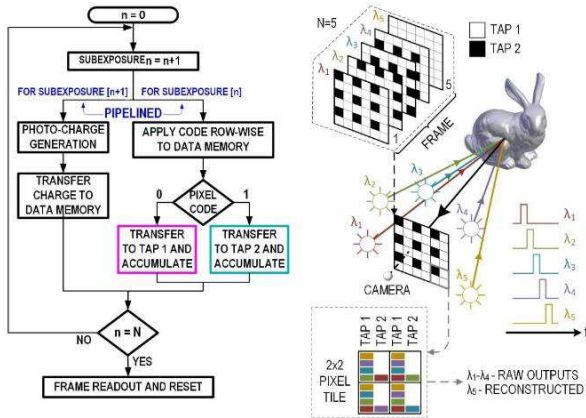
Fig. 1: The flow chart for the data-memory pixel (DMP) (left) and multispectral imaging setup (right).

Figure 1 (right) depicts the MSI camera setup which uses 5 LED lights sources, each with a unique wavelength. The 5 LEDs are sequentially turned on during the 5 respective subexposures. Synchronously with the LEDs, 5 code matrices, organized in 2x2-pixel tiles repeated over the entire pixel array, are submitted

to the sensor. The sorted photogenerated charges are accumulated during the 5 subexposures and are read out once at the end of the frame as two images. From the 8 taps of each 2x2-pixel tile, 5 images at the 5 different wavelengths are then extracted by demultiplexing [9]. The number of wavelengths is flexible and can be easily extended using a larger pixel tile, at the cost of a modest reduction in the spatial resolution.

III. SENSOR ARCHITECTURE

The DMP schematic and timing diagrams are shown in Figures 2 and 3, respectively. For each subexposure, the photogenerated light is first globally buffered on the pinned storage diode (SD), which acts as the data memory, before being sorted between taps 1 and 2, row-by-row. The exposure on the pinned photodiode (PPD) and the sorting of the SD charge are done in a pipelined fashion as desrcribed in section II.

Figure 4 (top) depicts the global exposure period, where the photogenerated charge collected at the PPD is transferred to the SD, by turning on the transfer gate by the global signal TG_GLOBAL .
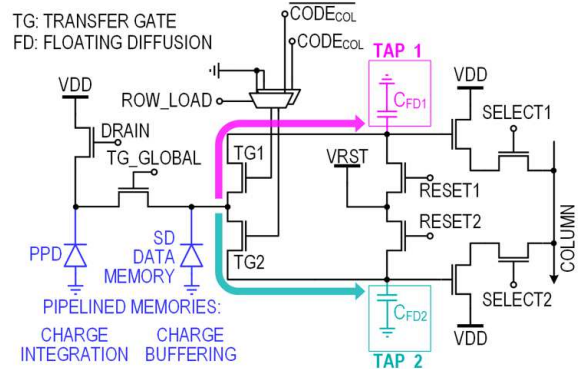

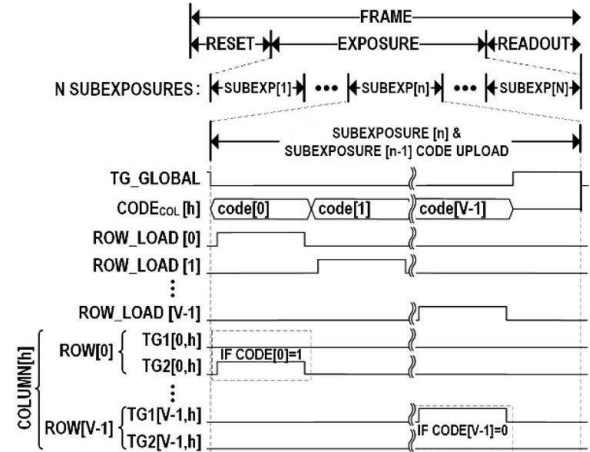Fig. 2: Schematic of the data-memory pixel (DMP).

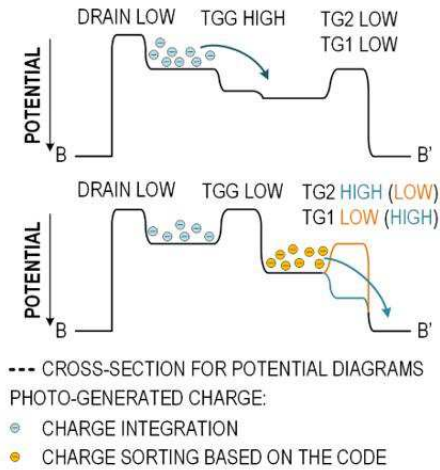Fig. 3: Timing diagram for the data-memory pixel (DMP).

Fig. 4: Potential diagrams for the global (top) and coded (bottom) transfer of the photogenerated charge in data-memory pixel (DMP).

Fig. 5: Data-memory pixel (DMP) simplified layout visualization.

Figure 4 (bottom) shows the coded integration of photogenerated charges. During each subexposure, rows are accessed one after another by activating their respective ROW_LOAD terminals, then the exposure code is applied. The coded-exposure operation for that subexposure is completed with the charges from SD being transferred to taps 1 or 2 for codes 0 and 1, respectively. As a result, the photogenerated charges across all subexposures of a frame are selectively integrated over the two taps according to the per-pixel code sequence and are read out once at the end of the frame as two images. An aggregated code transfer rate of 4Gbps and charge sorting transfer time of 1µs yield a pixel code rate of 270MHz, corresponding to approximately 2700 exposures per second at 312 x 320 sensor resolution.

The compact 7µm pixel layout is shown in Figure 5. A key challenge in pixel design for CEP image sensors is the area and time overhead due to the in-pixel exposure control circuits. Most existing CEP image sensors [3-6] employ a pixel architecture we refer to as the code-memory pixel (CMP) where an in-pixel memory to store the exposure code is used, which has significant area penalty resulting in a larger pixel. In this work, a two-tap pixel 2.5x smaller than the state of the art 2-tap CMP pixel [5] is achieved. The DMP architecture eliminates the need for in-pixel storage of the exposure code or, in fact, for any PMOS transistors.
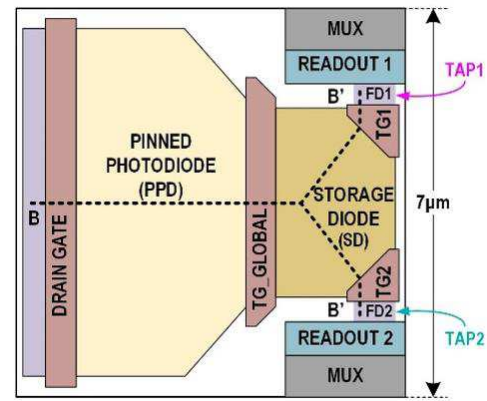
The top-level architecture of the sensor is depicted in Figure 6. The codes are sent to the sensor at a 4 Gbps rate and deserialized before being sent to 312 columns of pixels. This operation is done in a row-by-row fashion, taking advantage of the pixel's high-speed code transfer. The outputs of the pixels are made available by the horizontal readout scanner and serialized using 3 output channels. Figure 7 depicts the chip micrograph (left) and the camera prototype board (right). A compact FPGA board featuring Xilinx Artix 7 is used to interface with the sensor chip, providing input control signals and reading out the image data.
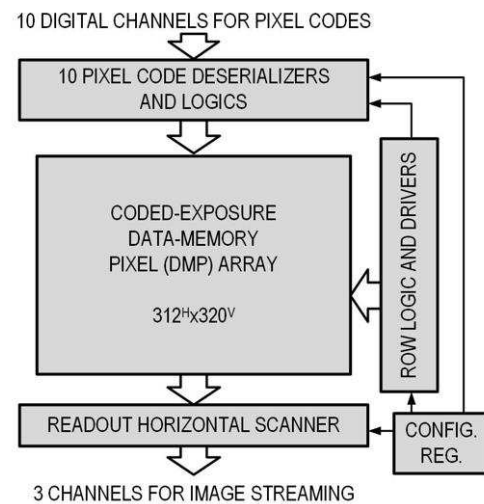

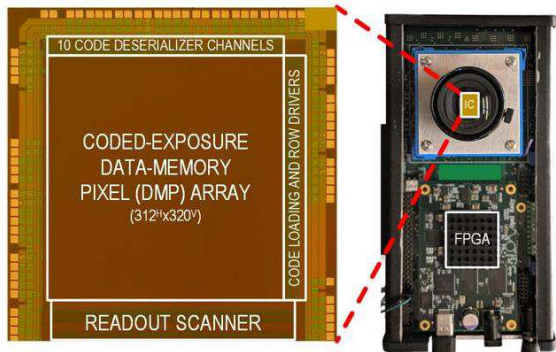Fig. 6: Top-level architecture of the image sensor.

Fig. 7: Chip micrograph (left) and camera prototype (right).

IV. EXPERIMENTAL RESULTS

The presented CEP image sensor has been employed to produce snapshot multispectral images, without using on-chip or off-chip color filters. The experimental procedure utilized the method depicted in Figure 1 (right) and described in Section II. Figure 8 depicts the imaged object - a colorful blanket (left) and the experimentally obtained output images of the camera for 5 different wavelengths (640nm to 940nm) demultiplexed from the two tap images of a single frame, at 30fps. A reconstructed RGB color image is also depicted (bottom, right). The images for the blue, green, and red wavelengths contain information from the visible light spectrum used to reconstruct the RGB color image. The NIR wavelengths are useful in making it possible to understand material surface texture and other properties.
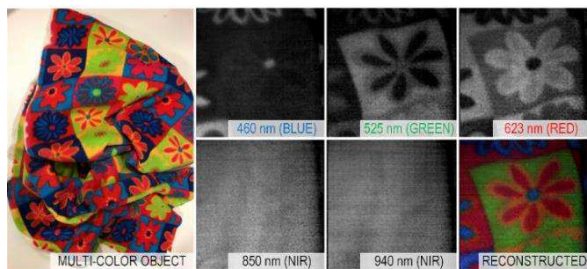
Fig. 8: Multispectral imaging experimental results at 30fps video rate. The scene on the left is captured by a conventional camera.

Table 1 summarizes the advantages of this work as compared to other CEP image sensors. The sensor achieves a 7µm pixel pitch and the maximum rate of 2700 coded subexposures per second. This subexposure rate can be used to increase the sensor frame rate when used with a small number of bands, or to trade the frame rate for a larger number of spectral bands by using larger tile sizes.

	THIS WORK	[3] UBC JSSC 2022	[4] Canon ISSCC 2022	[5] Toronto ISSCC 2019
PIXEL				
TECHNOLOGY [nm]	110 CIS	130 CIS	90 SPAD/ 40 CMOS	110 CIS
PINNED PHOTODIODE	YES	NO (PG)	NO (SPAD)	YES
PIXEL PITCH [µm]	7	12.6	9.5	11.2
FILL FACTOR [%]	38.5	38.7	~100	45.3
NUMBER OF TAPS	2	1	1	2
TAP CONTRAST[1]	90	N/A	N/A	99.5
ARCHITECTURE				
PIXEL COUNT [HxV]	312 x 320	192x192	**960 x 960**	244 x 162
FRAME RATE [fps]	30	30	**90**	25
POWER [mW]	54[3]	31.5	330	34.4
POWER FoM [nJ/frame·pixel]	18	28.5	**4**	34
SYSTEM				
IN-PIXEL CODE MEMORY	NO	YES (SRAM)	YES (SRAM)	YES (2 LATCHES)
IN-PIXEL DATA MEMORY	YES (CHARGE)	NO	YES (SRAM)	NO
SUBEXPOSURE RATE [kHz]	2.7	23	**370**	.18
PIXEL CODE-RATE [MHz]	270	850	340	7.1
ARBITRARY CODE / ROI[2]	YES/YES	YES/YES	NO/--	YES/YES
FRAME-CODE SHUTTER	GLOBAL/ROLLING	GLOBAL/ROLLING	GLOBAL	GLOBAL
IMAGING APPLICATIONS	Multispectral imaging	Compressive depth sensing	HDR imaging	1 Structured-light 2 Photometric stereo

1: Also known as Extinction Ratio   N/A: Not Applicable; --: Not Available
2: ROI: region of interest   Bold font denotes the best performance
3: without ADC power

Table 1: Comparison Table.

V. CONCLUSIONS

We have demonstrated flexible spectrally-scanning snapshot multispectral imaging on a dual-tap coded-exposure-pixel CMOS image sensor. Using the data-memory coded-exposure pixel, scene images for 5 wavelengths in the visible light and NIR spectrum obtained at are obtained within a single frame, at 30 frames per second, without using a color filter array.

REFERENCES

[1] J.-I. Park, et al., "Multispectral Imaging Using Multiplexed Illumination," ICCV, 2007.

[2] G. Wan, et al., "CMOS Image Sensors with Multi-Bucket Pixels for Computational Photography," JSSC, 2012.

[3] Y. Luo, et al., "A 30-fps 192x 192 CMOS Image Sensor with Per-Frame Spatial-Temporal Coded Exposure for Compressive Focal-Stack Depth Sensing," JSSC, 2022.

[4] Y. Ota, et al., "A 0.37W 143dB-Dynamic-Range 1Mpixel Backside-Illuminated Charge-Focusing SPAD Image Sensor with Pixel-Wise Exposure Control and Adaptive Clocked Recharging," ISSCC, 2022.

[5] N. Sarhangnejad, et al., "Dual-Tap Pipelined-Code-Memory Coded-Exposure-Pixel CMOS Image Sensor for Multi-Exposure Single-Frame Computational Imaging," ISSCC, 2019.

[6] H. Ke, et al, "Extending Image Sensor Dynamic Range by Scene-aware Pixelwise-adaptive Coded Exposure," IISW, 2019.

[7] L. Wang, et al., "High-speed hyperspectral video acquisition with a dual-camera architecture," CVPR, 2015

[8] C. Xun, et al., "Computational Snapshot Multispectral Cameras: Toward dynamic capture of the spectral world," IEEE Signal Processing Magazine, 2016-09

[9] M. Wei, et al., "Coded two-bucket cameras for computer vision," ECCV, 2018.

A SPAD-based linear sensor with in-pixel temporal pattern detection for interference and background rejection with smart readout scheme

Alessandro Tontini*†, Leonardo Gasparini†, Roberto Passerone*
*University of Trento, via Sommarive 9, 38123 Trento, Italy
†Fondazione Bruno Kessler, via Sommarive 18, 38123 Trento, Italy
Email: tontini@fbk.eu; gasparini@fbk.eu; roberto.passerone@unitn.it

Abstract—In this work, a 1x64 pixel SPAD-based linear sensor for direct time-of-flight (d-ToF) applications with real-time in-pixel interference and background rejection is presented. Each pixel is composed by 4 SPADs with passive quenching, a digital logic circuit to exploit photon temporal coincidence with a threshold of up to 3 photons for background rejection, a finite state machine for the detection of temporal laser patterns for the rejection of interfering signals generated by other similar devices and a 16-b time-to-digital converter with 150 ps timing resolution that can be repurposed for intensity measurements. The sensor implements a smart readout scheme capable to output only pixels with meaningful data, i.e., detection events that have been validated by the photon temporal coincidence circuit and/or the laser pattern detection circuit.

Index Terms—Single Photon Avalanche Diode (SPAD), Light Detection And Ranging (LiDAR), direct time-of-flight (d-ToF).

I. INTRODUCTION AND RELATED WORK

3D sensing-capable devices are widespread in several fields, from the industry to the consumer market. Among the many techniques that provide depth information, direct time-of-flight (d-ToF) with SPADs implemented in CMOS technology proved to be one of the most promising, thanks to the integration capability of the most advanced process nodes. Among the many challenges that SPAD-based CMOS d-ToF sensors have to face, dealing with background light and interference from similar devices are currently the limiting factors.

Background light rejection has been extensively studied and several techniques proved to be effective, such as photon temporal coincidence [1], time-gated acquisitions [2] or, for long-distance targets, the detection of the last incoming photons [3]. On the other hand, with the rapid diffusion of such devices, the need to deploy solutions for mutual interference suppression is capturing the attention of researchers even more than background light, especially in autonomous-driving scenarios.

Notable works provided effective solutions to reduce the disturbance effect from unwanted laser sources. Ximenes et al. [4] implemented a phase-shift keying (PSK) modulation on the emitted laser pulse, thus spreading interference sources below the level of the signal of interest. Seo et al. [5] exploit interference suppression by means of the emission of two laser pulses, whose timing signature is recognized and used to reject any other source with a different temporal pattern.

Both approaches demonstrated their effectiveness in reducing the effect of interference. With a PSK modulation of the emitted laser pulse [4], interference is not eliminated, but only attenuated, where the efficiency of attenuation depends on the number of phase shifts of the modulation. On the other hand, it has the advantage of requiring only one laser pulse, as opposed to the second approach, where two laser pulses with a known timing signature are emitted [5]. This approach, however, allows for a complete cancellation of interfering sources, as unknown detections are eliminated rather than attenuated. In the work from Seo et al. [5], this technique is exploited in post-processing over the collected histogram of timestamps with promising results.

In this work, we present a 1x64 pixel SPAD-based linear sensor embedding in-pixel background and interference rejection, with a smart readout scheme capable of selecting only pixels containing validated data to improve the sensor rate of operation. The interference rejection is based on the emission of two laser pulses with a known timing signature as in [5], but in this work it is implemented directly on a pixel basis in a compact form and operates in real time, thus no post-processing is required on the histogram of timestamps. The benefit of an active, in-pixel interference rejection is twofold. First, also background light can be rejected, resulting in an increased signal to background ratio in the final histogram. Second, power consumption is reduced as the TDC is activated only when two photon detections occur within the expected time frame, resulting also more robust against pile-up distortion, as the probability of saturating the TDC channel is reduced.

The paper is organized as follows. A detailed description of the array architecture is provided in Section II, focusing in particular on the pixel architecture (Section II-A) and on the readout scheme (Section II-B). Characterization results are reported in Section III, while conclusions and perspectives for future improvements are discussed in Section IV.

II. ARCHITECTURE

In this section, we describe the sensor in detail, focusing on the pixel architecture and on the readout scheme.

A. Pixel architecture

The pixel, designed in a 110 nm 4M CIS technology, is composed by 4 SPADs arranged as a mini digital silicon photomultiplier. Each SPAD is passively quenched by 2 thick-oxide transistors to recharge the SPAD and properly clamp the voltage to 1.2 V to be compliant with the following circuitry. Each SPAD is paired with a monostable circuit to create a temporal window for the coincidence detection circuit, which is realized in pure digital logic. A threshold of N=1/4, N=2/4 and N=3/4 events can be selected via SPI programming. The output from the coincidence detection is fed into the measurement control circuit, which implements a finite state machine for the detection of the laser signature. The laser temporal signature can be set with 4-bit granularity, i.e., up to 16 combinations are possible. The TDC is based on a fine-coarse architecture, where the coarse timing measurement is given by an 8-bit counter with 100 MHz clock and the fine timing by a ring oscillator with 150 ps timing resolution. The TDC 8-bit counter can be repurposed to count the number of detected photons for intensity measurements. When the sensor is operated to recognize the laser pattern to reject interference, the TDC is triggered only when the second laser pulse is correctly detected, reducing unnecessary power consumption.

The measurement control circuit generates a *VALID* flag whenever a photon-detection event occurs. If the laser pattern detection feature is disabled, a *VALID* flag is generated by the first incoming event, which can be either the first detected photon (if no coincidence threshold is applied) or the first 2 or 3 photons detected within the coincidence window generated by the monostable circuit. The *VALID* flag is needed by the smart readout scheme to optimize the bandwidth by reading only pixels with validated data.

The chip micrograph, array architecture and pixel block diagram are shown in Figure 1.

B. Readout architecture

With the capability to exploit both the photon coincidence and laser signature detection, a reduction in the amount of generated data is expected. For this reason, a classical scheme where the entire array is read out is not optimized, as pixels with either non-validated data or background data are anyway considered, resulting in an increased readout time, negatively affecting the sensor frame rate. For this reason, we implemented a dedicated readout circuit which is able to output only pixels with meaningful data, i.e., detection events that have been validated by the photon coincidence circuit and/or the finite state machine for the detection of the temporal laser pattern. The readout process comprises two phases: the first phase is meant to transfer 1 bit per pixel to inform the controller FPGA about which pixels contain valid data. In the second phase, only pixels with validated data are read out, thus suppressing zeros. The first readout phase is therefore only needed for the controller FPGA to associate each data values with the pixels that generated it, with a minimum overhead of only 1 bit per pixel.
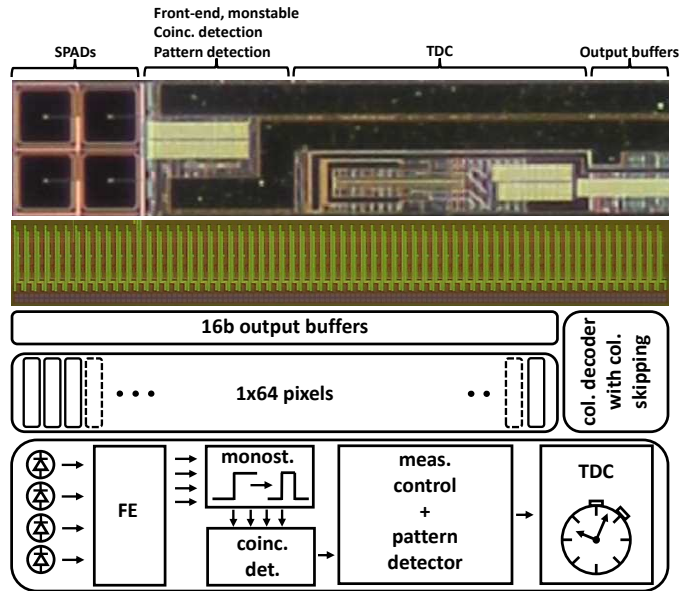

Fig. 1. Chip micrograph, array architecture and pixel block diagram. The array of pixels is implemented in a 110 nm 4M CIS technology within a multi-project chip. Due to the reuse of the TDC from a previous project, pixel size is not optimized, resulting in a final pitch of 40 x 180 μm^2. The in-pixel measurement control block, with laser signature detection capability, has an area occupation of 28 x 14 μm^2. By considering the device from Manuzzato et al., [3], which is realized in the same technology node with a pixel pitch of 48 x 48 μm^2, the occupation of this block takes 17% of the total pixel area, thus allowing its integration also in a 2D array.

The proposed readout scheme well matches with the in-pixel implementation of the laser signature detection, as the generation of data from the sensor can be highly reduced with such a strong filtering. Consequently, the benefit in terms of performance is twofold: on one side, the reliability of the timestamp detected with the in-pixel finite state machine for the laser pattern detection is increased, and on the other side the reduced amount of data (mainly due to the filtered events) results in a decrease of the required readout time, with benefits in terms of frame rate and power consumption.

III. CHARACTERIZATION

In this section, characterization results focusing on the in-pixel laser pattern detection and readout circuit performance are shown.

A. In-pixel laser pattern detection characterization

The in-pixel laser pattern detection feature has been first characterized on a single pixel basis using only the coarse TDC (100 MHz counter) information by means of two low-power picosecond pulsed lasers. The first laser was meant to be the signal source, while the second was used as an interferer. A picture describing the experimental setup is shown in Figure 2. The sensor is operated in the presence of background light, without any optical bandpass filter in front of the detector and with a coincidence threshold of N=2 photons. The results of the first characterization are shown in Figure 3, demonstrating that an interference with a histogram peak 18.5 dB higher than
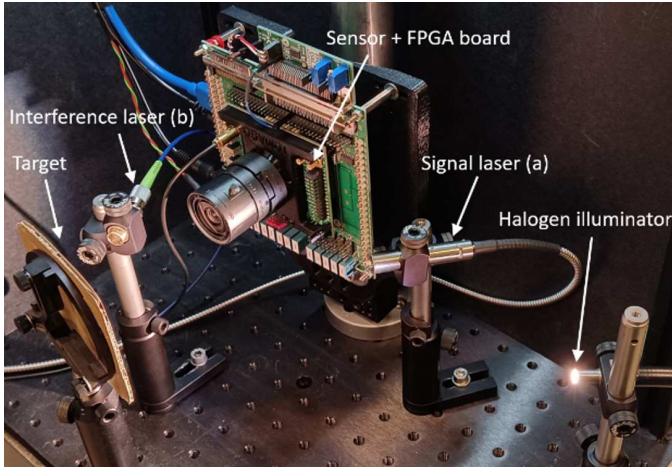
Fig. 2. Picture of the experimental setup to test the in-pixel interference-rejection capability with the signal laser (a) and the interfering laser (b). Each laser has been set with its own timing signature: for the signal of interest, the two laser pulses are separated 80 ns from each other, while for the interfering signal the pulses separation is 90 ns.
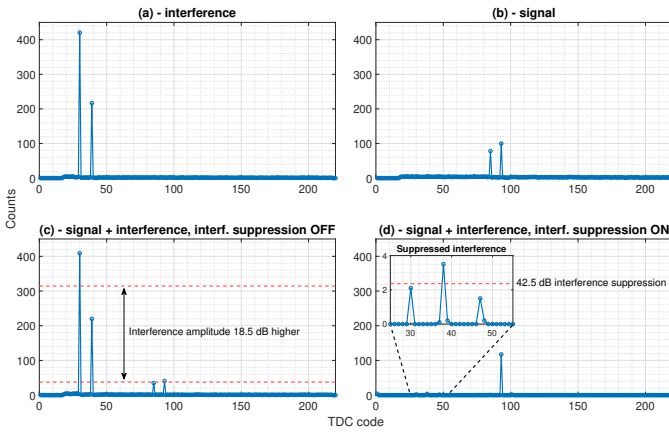

Fig. 3. First experimental validation of the interference-rejection capability of the device. Histograms (a) and (b) show the contribution of interference (a) and signal (b) alone, while in (c) the joint effect of the two sources is shown (with laser signature detection disabled). To stress the interference-rejection capability, the interference histogram peak is 18.5 dB higher in amplitude than the peak of the signal of interest. When the in-pixel interference-rejection is enabled (d), the interfering signal is almost completely suppressed by 42.5 dB, whereas the signal of interest gains 10 dB with respect to case (c), enabling the possibility to build an interference-free histogram directly from pixel data.

the signal of interest is almost completely eliminated, with a suppression of 42.5 dB. On the other hand, the signal of interest gained 10 dB with respect to the case where no laser pattern matching detection was applied, demonstrating that the in-pixel active detection of laser temporal patterns also helps in reducing uncorrelated background light.

Then, the entire array has been characterized with the 3D measurement of a scene profile, using the fine resolution given by the 150 ps in-pixel ring-oscillator and employing two 25 W 905 nm lasers for the emission of the temporal laser pattern. A third laser was pointed toward a portion of the scene, to emulate the interference from a second device. Results are shown in Figure 4.
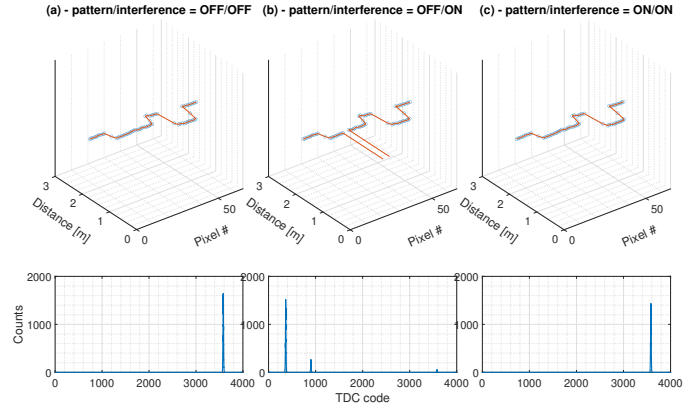
Fig. 4. Measurement of the profile of a scene composed by 2 boxes with different size and distance (range 0-2m). The first measurement (a) is a reference obtained without any interfering signal and with pattern detection disabled. In the second measurement (b), the interfering laser is pointed toward the first box with an earlier timing with respect to the signal laser, resulting in a complete loss of information from the illuminated portion of the target. In the third measurement (c), the in-pixel laser pattern detection was enabled allowing for a complete recovering of the lost information, as the interfering laser timestamps are actively discarded in favor of the signal laser timestamps, which have the correct timing. For each measurement, the histogram of one pixel under the interfering portion of the target is shown.

With a third measurement we targeted the background-rejection capability of the sensor by focusing on the combination between the photon coincidence technique and the in-pixel laser pattern detection. The sensor was operated without optical bandpass filter and a 180 W halogen illuminator was used to flood the scene with background photons. Without any background/interference rejection features, it was impossible to reconstruct the 3D profile of the scene. The 3D information was recovered completely with the photon coincidence technique (operated with a threshold of N=2/4 photons), but still several background events are allowed to trigger a coincidence, increasing the amount of data generated by the sensor and thus the required readout time. By enabling the laser pattern detection on top of the photon coincidence, it was anyway possible to reconstruct the 3D profile of the scene and at the same time dramatically reduce the amount of false triggers, with a reduction of data of $\simeq 98\%$ with respect to the previous case. By considering the laser peak to background ratio of the histogram, the combination of the laser pattern matching on top of the photon coincidence detection allowed to gain $\simeq 23$ dB, thus increasing the robustness of the measurement. Results are shown in Figure 5.

B. Readout performance assessment

The readout architecture has been characterized in two ways. The first characterization aimed at showing the capability of the sensor to output only pixels with meaningful data. For this measurement, a target (flat panel) was illuminated completely by background light and a collimated laser source was used to illuminate only a portion of it, resulting in only 2 pixels of the array to be shined with laser photons. First, no background-rejection technique was applied nor laser pattern
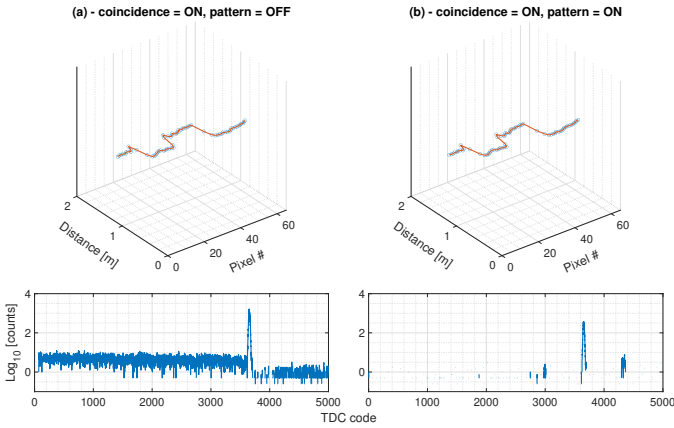
Fig. 5. Measurement of the 3D profile of a scene with high background light. In (a), only the photon coincidence technique was enabled, with a threshold of N=2/4 photons. In (b), the laser pattern detection was enabled on top of the photon coincidence, resulting in ≃ 98% data reduction with respect to (a). The laser peak to background ratio is ≃ 50 dB in (a) and increased up to ≃ 73 dB in (b).
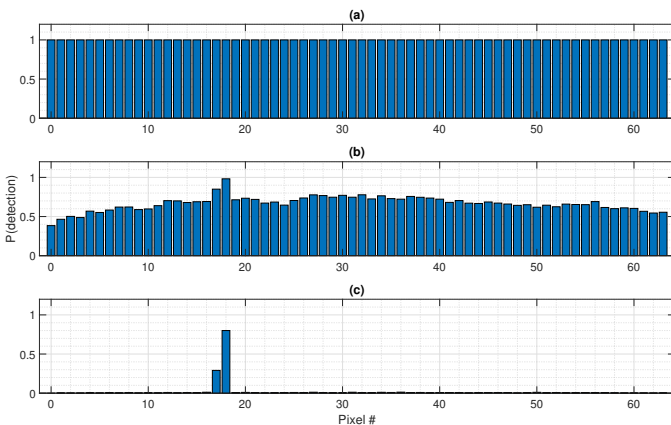

Fig. 6. Measured per-pixel activity (in terms of probability of detection) with the proposed readout scheme. With a minimum 1-bit overhead per pixel, it is possible to output only pixels with validated data. In (a), the readout ratio is almost 100% as no data filtering technique was applied. In (b), the enabling of the photon coincidence technique allows to reduce the readout ratio down to 66%, with a visible readout peak activity over pixels number 17 and 18, as they coincide with the reflected laser footprint. In (c), the additional enabling of the in-pixel laser pattern matching detection allows to recover the information from the only two illuminated pixels, further reducing the readout ratio down to a minimum of 2.25%.

matching detection, resulting in a probability of detection of almost 1 for all pixels. Then, photon coincidence detection was enabled with a threshold of N=2 photons, and a general reduction of the probability of detection is observed, but still allowing background-only pixels to be triggered. In the last measurement, the in-pixel laser pattern detection was enabled on top of the photon coincidence, reducing the probability of detection on pixels not illuminated by the laser source to a negligible level, allowing to only output data from the subset of pixels illuminated by the laser. Results are shown in Figure 6.

In the second characterization, only background light was considered and the sensor internal frame rate was measured by controlling the intensity of the light source. Results are shown
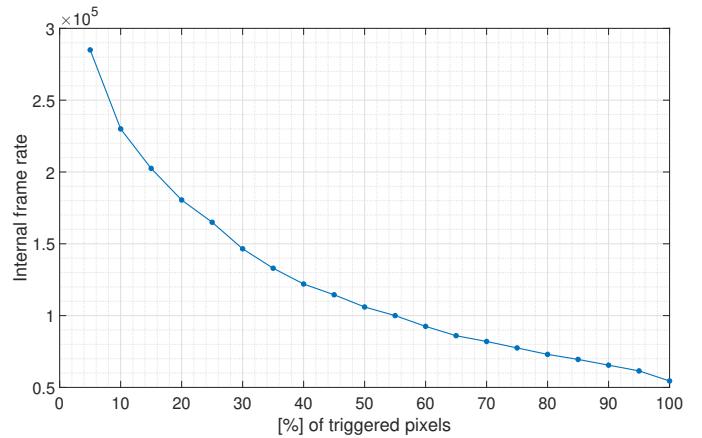

Fig. 7. Measured sensor internal frame rate (thus not considering PC elaboration time) for different number of triggered pixels, demonstrating the capability of the proposed readout scheme to adapt to the level of photon activity.

in Figure 7, showing the capability of the sensor to adapt to different levels of photon activity.

IV. CONCLUSION

In this work, a SPAD-based d-ToF linear sensor with the first in-pixel interference rejection capability has been demonstrated. In combination with the dedicated readout architecture, it enables the acquisition of interference-free data with an optimized, adaptive, readout time. The sensor capability to reject interference from other laser sources has been demonstrated in a laboratory setup, as well as the opportunity to use the interference rejection feature on top of the classical photon coincidence to reduce background light to a negligible level, resulting in a measured 42.5 dB interference suppression and up to 23 dB signal gain with a measured data compression ratio of ≃ 98%. The pixel pitch in its actual form factor is not optimized for the evolution into a 2D array and for such reason we envisage a further development with a more compact TDC to extend the proposed pixel architecture into a 2D imager array.

REFERENCES

[1] M. Perenzoni et al. A 64 × 64-pixels digital silicon photomultiplier direct TOF sensor with 100-Mphotons/s/pixel background rejection and imaging/altimeter mode with 0.14% precision up to 6 km for spacecraft navigation and landing. *IEEE Journal of Solid-State Circuits*, 2017.
[2] P. Padmanabhan et al. A 256×128 3D-stacked (45nm) SPAD FLASH LiDAR with 7-level coincidence detection and progressive gating for 100m range and 10klux background light. In *2021 IEEE International Solid-State Circuits Conference (ISSCC)*, 2021.
[3] E. Manuzzato et al. A 64 × 64-pixel flash LiDAR SPAD imager with distributed pixel-to-pixel correlation for background rejection, tunable automatic pixel sensitivity and first-last event detection strategies for space applications. In *2022 IEEE International Solid- State Circuits Conference (ISSCC)*, 2022.
[4] A. R. Ximenes et al. A 256×256 45/65nm 3d-stacked spad-based direct tof image sensor for lidar applications with optical polar modulation for up to 18.6db interference suppression. In *2018 IEEE International Solid - State Circuits Conference (ISSCC)*, 2018.
[5] H. Seo et al. Direct tof scanning lidar sensor with two-step multievent histogramming tdc and embedded interference filter. *IEEE Journal of Solid-State Circuits*, 2021.

Count-Free Histograms with Race Logic for Single-Photon LiDAR

Atul Ingle   David Maier

Portland State University

Abstract—**Low-power 3D perception is useful in a wide range of computer-vision applications. Thanks to the increasing availability of high-resolution single-photon avalanche diode (SPAD) arrays, single-photon LiDARs (SPLs) have emerged as a promising technology for 3D sensing. The conventional image formation model for an SPL involves capturing the time-varying light intensity—which we call the *transient distribution*—of a reflected laser pulse in the form of an *equi-width* (EW) histogram. Unfortunately, this approach leads to unmanageable data rates (∼gigabytes/second) with high-resolution arrays, severely limiting the applicability of SPLs in power- and bandwidth-constrained scenarios (e.g., mobile devices). We propose a radically different approach based on race logic processing to construct equi-depth histograms with variable bin widths. This method avoids storing high-resolution histogram counts, thereby reducing the bandwidth requirement while maintaining similar ranging accuracy. We show simulation results with bandwidth reduction of over $100\times$.**

Index Terms—**LiDAR, 3D imaging, SPAD, compression, race logic**

I. INTRODUCTION

Single-photon sensing is a promising technology for high-resolution 3D imaging. Low-power 3D perception is useful in a range of computer-vision applications, including industrial robotics, autonomous driving and augmented reality. Image sensors that can capture single photons, such as single-photon avalanche diodes (SPADs), are popular as detectors for such applications. High (kilo-to-megapixel) resolution SPAD arrays with additional data-processing embedded in the same hardware are increasingly available. However, their high sensitivity and speed is a double-edged sword: the data generated by such arrays greatly exceeds what can be reasonably processed or transferred in real time, limiting their applicability, especially where there are power and bandwidth constraints. A single-photon camera (SPC) captures depth information using the time-of-flight principle [4]. A laser illuminates the scene with a short light pulse. The corresponding camera pixel captures a stream of return events as photons arrive at different delays with respect to the pulse time. The aggregation of these delays over many laser cycles forms a transient distribution that we wish to capture. Traditional methods construct
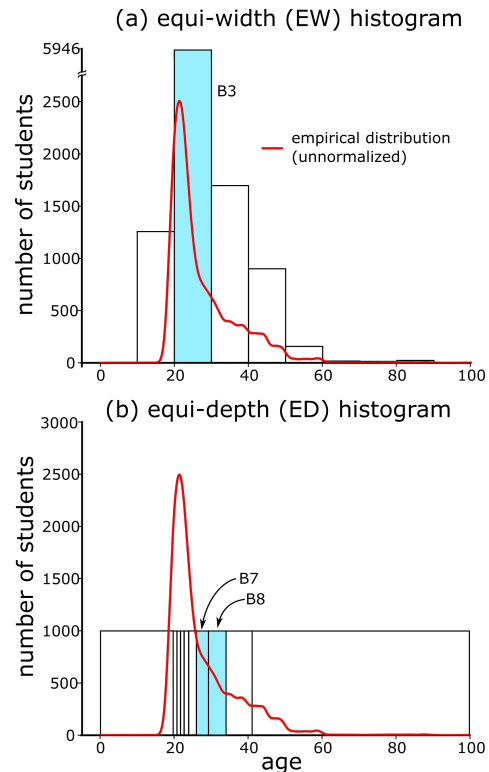
Fig. 1. **Advantage of equi-depth (ED) histogram over conventional equi-width (EW) histogram for peaky distributions.** This figures shows two types of histograms for student-age data from Dale et al. [2]: (a) A 10-bin EW histogram has a single bin B3 near the peak. Many bins are close to zero and provide no useful information about the peak location. (b) A 10-bin ED histogram with approximately 1000 students per bin reliably captures the shape of the peak. In this work we apply this intuition to SPL data: ED histograms adaptively cluster around the transient-distribution peak, providing accurate distance information with only a few ED bins.

histograms of 1000's of bins by storing event counts at different delays over multiple laser cycles. The peak of a histogram gives an estimate of the true distance of the scene point (by the relationship that speed of light × delay = twice the distance to the scene point).

II. KEY IDEAS

We want to estimate scene distances using an SPC with minimal power and data bandwidth. This goal rules out forming 1000-bin histograms on the image sensor, which requires large storage on chip and high bandwidth during readout of bin counts. Conventional histogram-based methods are especially inefficient with strong background illumination, as
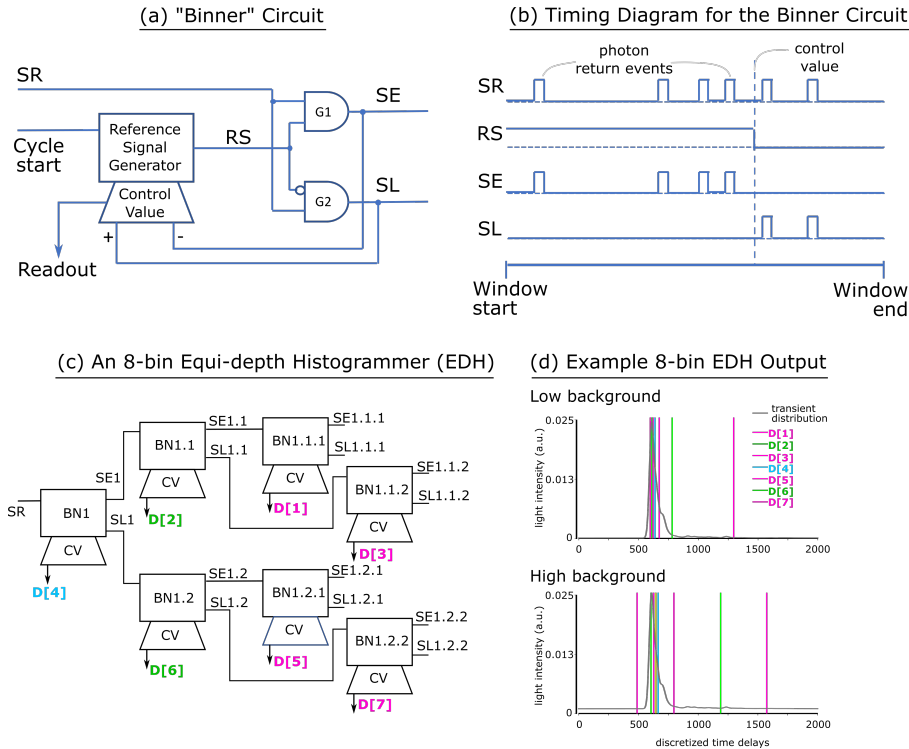
Fig. 2. **The proposed binner circuit and equi-depth histogrammer (EDH).** (a) The binner circuit splits the incoming photon stream (SR) into an early stream (SE) and a late stream (SL) depending on transition point of a reference signal (RS) generated from a control value. (b) In this example, there are more photons in the early stream than the late stream, so the control value will decrease for the subsequent laser cycle, thus moving the transition point of RS earlier. The control value eventually settles close to the overall median. (c) An 8-bin ED histogram can be captured using a collection of 7 binners arranged in a 3-level binary tree. A binner at one stage feeds streams of early and late photon events to two binners at the next stage in the tree. (d) This example shows a transient distribution and the simulated results of an 8-bin EDH for low and high background levels. Notice that a majority of the bins cluster around the true peak location. The location of the narrowest bin provides a reliable estimate of scene distance.

most bins count background photons. We offer a radically different approach for time-of-flight imaging compatible with multiple detector and illumination schemes. It has two key elements: constructing equi-depth (ED) histograms rather than the traditional equi-width (EW) histograms, thereby using many fewer bins to approximate the transient distribution and using race logic to process information in the "delay domain", avoiding conversion of return events to digital timestamps.

Equi-Depth Histograms: ED histograms have *variable-width* bins each with approximately the same number of items. Fig. 1 shows a conceptual example of 10-bin EW and ED histograms for a dataset of 10,000 Canadian college students [2]. The majority of this population is in the bin labeled "B3" of the EW histogram, while some other bins are empty, contributing little information. In contrast, a 10-bin ED histogram for the same distribution does a better job of capturing the shape of the peak. In case of an SPL, ED histograms can provide reliable estimates of scene distance with very few bins that cluster around the peak of the distribution.

Race Logic: Race logic [5] is a novel approach to computation in which values are represented as time delays, rather than as analog or digital quantities.
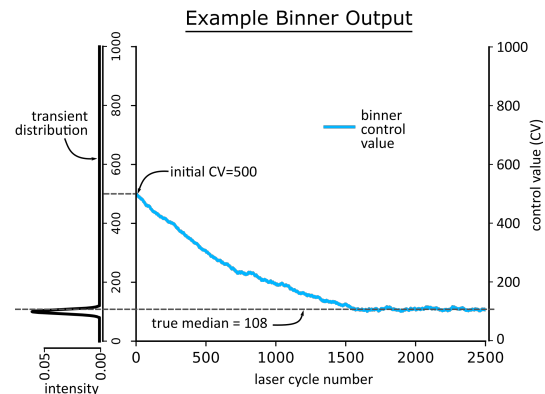

Fig. 3. **Example trajectory of the control value of a binner.** We show a binner's CV over multiple laser cycles. The true peak is at 100, signal strength is 1.0 and signal-to-background ratio is 1.0.

Race logic is well suited to SPAD data processing, as the photon arrivals are already in the delay domain. Working with information in this form avoids the energy required to convert return events to digital timestamps, which most existing methods need.

III. METHODS

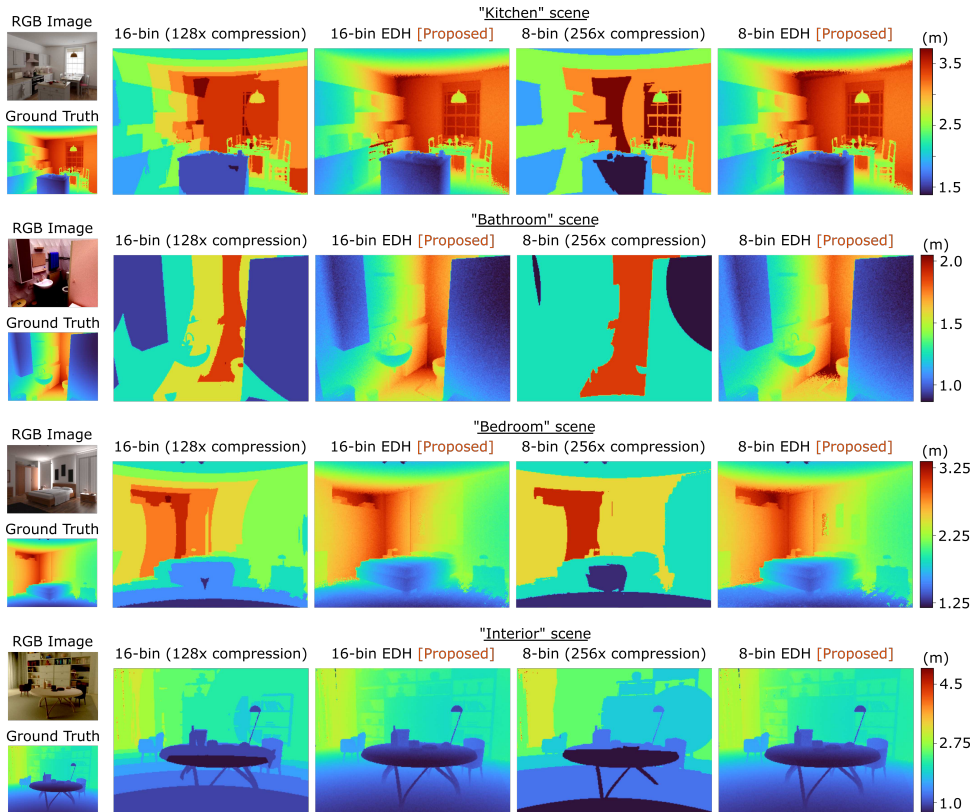Fig. 2(a) shows our proposed binner circuit. "Early" or "late" photons relative to the current estimate (called

Fig. 4. **Simulated results using a transient rendering dataset [3].** This figure shows results for "Kitchen", "Bathroom", "Bedroom" and "Interior" scenes from a simulated transient-rendering dataset. RGB and ground-truth distance maps are shown alongside distance map reconstructions using conventional 8-bin and 16-bin coarse EW histograms. Observe the lack of details and increased quantization artifacts in distance maps computed using the compressed coarse EW histogram method. In contrast, our method reliably captures scene distances with as few as 8 ED bins while achieving over $100\times$ compression.

the "control value" (CV)) move that estimate earlier or later, respectively. Photon-timing histograms are never explicitly stored in memory (unlike equi-width-histogram approaches that remember the timestamp of every photon received). Intuitively, the control value must settle at a location where equal number of photons arrive in the early and late streams. By definition, this point is the median of the transient distribution. Note, however, that the binner does not converge deterministically to the median because photon arrivals are random and the binner does not maintain the full history of photon arrivals. Fig. 3 shows the CV of a simulated binner over multiple cycles, for a synthetic transient distribution with a single, narrow peak, along with the true median of the distribution.

The median estimate will often lie close to the peak in the transient distribution, but can deviate significantly, if there is significant background light or the peak is asymmetric. To deal with such divergence, we cascade multiple binners stages to form an equi-depth histogrammer (EDH). The collective CVs of these binners represent the ED histogram bin boundaries. A binner at one stage feeds streams of early and late return events to two binners at the subsequent stage, each which in turn subdivides the corresponding bin from the previous stage. For example, a three-stage

EDH shown in Fig. 2(c) will compute the boundaries of a 8-bin ED histogram. With multiple bins, the bin boundaries cluster around the transient-distribution peaks. Fig. 2(d) shows an example simulated output for a transient distribution with a single peak and some background illumination. Having adaptive bin boundaries helps an EDH cope with background light. The lower part of Fig. 2(d) uses the same simulated transient distribution as the upper part, but with more background light. We see a couple more bins away from the peak, which essentially "absorb" the background photons. However, there are still sufficient bins remaining to capture laser peak well. Since EDH bin boundaries cluster around the peak of the transient distribution, scene distances can be estimated by locating the narrowest bin. We are evaluating two methods for estimating the transient peak from an ED histogram. The simpler method just returns the midpoint of the narrowest bin. That estimate can be inaccurate when two bins split the peak. Thus, we also have a more sophisticated method that fits a curve to the bins in the neighborhood of the narrowest bin, then uses the argmax of that curve as the peak estimate.

IV. RESULTS

We evaluate the performance of an EDH using a transient rendering dataset [3]. We simulate photon-

In-Pixel Binner Implementations (Speculative)

(a) Analog Implementation
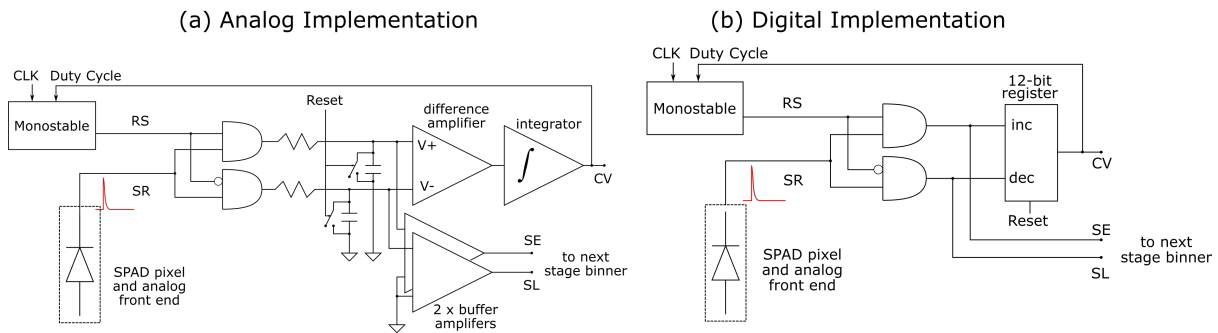

(b) Digital Implementation

Fig. 5. **Speculative pixel architectures for an in-pixel binner implementation.** (a) An analog implementation accumulates the early and late stream signals as charges on two capacitors that are compared to increase/decrease the control value. (b) In the digital implementation, an up/down register holds the control value.

return streams using the pixelwise ground-truth transient distributions from this dataset. Our simulation accounts for the effect of signal and ambient light strength, different scene albedoes, and sensor-noise sources including shot noise and dark counts.

Fig. 4 shows simulated transient-rendering results using 8- and 16-bin EDHs for four different scenes. Observe that in the "kitchen" scene (first row), our method preserves distance gradients of flat surfaces (such as the walls of the room) that appear quantized with a coarse 8- or 16-bin EWH. In the "bedroom" scene, an EDH reliably captures fine distance details such as the bed-frame with as few as 8 bins. In the "interior" scene, a coarse EWH loses details such as small objects on the table and in the shelf in the background. Our method preserves these distance details with as few as 8-bins. Unlike coarse EW histograms, our method avoids strong quantization artifacts and preserves details in the distance maps while reducing the data rate by $> 100\times$

V. Conclusion and Future Directions

We have presented a method for distance imaging that is a marked departure from current approaches. Instead of capturing the transient distribution by storing photon count histograms, we instead use a binner element that maintains and adjusts the distribution-median estimate at every laser-pulse cycle. Using a cascade of binners, we can produce equi-depth histogrammers that robustly capture peaks in the distribution. Our approach requires many fewer bins than with the equi-width approach, thus reducing memory and bandwidth requirements. Moreover, part or all of a binner element can operate in the delay domain using race logic, avoiding time-to-digital conversion, thus lowering circuitry and energy requirements.

Current single-photon cameras are severely bandwidth-constrained due to the requirement of reading out individual photon timestamps at extremely high (mega-to-gigahertz) rates. ED histograms captured using race logic processing have the potential to reduce this bandwidth requirement

by orders of magnitude, enabling future SPAD arrays to be scaled to higher spatial resolution. Lower bandwidth requirements can also provide power savings by requiring fewer bits to be moved off sensor, and may simplify in-pixel circuitry by avoiding the need to construct and store high-bit-depth histograms on-sensor.

Our current work is proceeding along two tracks, method refinement and hardware prototyping.

Method Refinement: We are exploring variations and extensions of the EDH approach that improve convergence rates and accuracy, as well as conducting further simulations and analysis to understand the effects of signal strength, background light and peak position on binner and EDH behavior.

Hardware prototyping: We are currently prototyping a binner circuit using a SPAD array and associated FPGA [1], and plan to extend that to a full EDH. Based on what we learn from that effort, we will explore designs that integrate the binner onto the acutal detector device. Fig 5 shows two (speculative) designs with a binner implemented in-pixel.

References

[1] Claudio Bruschini, Samuel Burri, Ermanno Bernasconi, Tommaso Milanese, Arin C Ulku, Harald Homulle, and Edoardo Charbon. Linospad2: a 512x1 linear spad camera with system-level 135-ps sptr and a reconfigurable computational engine for time-resolved single-photon imaging. In *Quantum Sensing and Nano Electronics and Photonics XIX*, volume 12430, pages 126–135. SPIE, 2023. 4

[2] Meghan Dale. Trends in the Age Composition of College and University Students and Graduates (Archived). *Education Matters: Insights on Education, Learning and Training in Canada*, (5), Dec 2010. Online (accessed July 21, 2022) https://www150.statcan.gc.ca/n1/en/pub/81-004-x/2010005/article/11386-eng.pdf. 1, 2

[3] Felipe Gutierrez-Barragan, Huaijin Chen, Mohit Gupta, Andreas Velten, and Jinwei Gu. itof2dtof: A robust and flexible representation for data-driven time-of-flight imaging. *IEEE Transactions on Computational Imaging*, 7:1205–1214, 2021. 3

[4] Istvan Gyongy, Neale AW Dutton, and Robert K Henderson. Direct time-of-flight single-photon imaging. *IEEE Transactions on Electron Devices*, 69(6):2794–2805, 2021. 1

[5] Advait Madhavan, Timothy Sherwood, and Dmitri Strukov. Race logic: A hardware acceleration for dynamic programming algorithms. *ACM SIGARCH Computer Architecture News*, 42(3):517–528, 2014. 2

A Study on Two Step Reset LOFIC Pixel to Reduce SNR Gap

Kazuki Tatsuta*, Ai Otani*, Shunsuke Okura*,
Ken Miyauchi†, Han Sangman†, Hideki Owada†, and Isao Takayanagi†
*Research Organization of Science and Engineering Ritsumeikan University,
1-1-1 Noji-Higashi, Kusatsu, Shiga, 525-8577, Japan
Email: ri0091ff@ed.ritsumei.ac.jp, Phone: +81-77-599-3149
†Brillnics Japan Inc., 6-21-12 Minami-Oi, Shinagawa-ku, Tokyo, 140-0013 Japan,

I. Introduction

High dynamic range (HDR) CMOS image sensors (CISs) are expected for machine vision purpose under extreme illumination conditions such as outdoor. An image sensor that has lateral overflow integration capacitor (LOFIC) [1] is one solution to realize the HDR-CIS. The LOFIC pixel is composed of typical 4 transistors, an overflow capacitor (CS) and a switching gate (SG). During high conversion gain (HCG) mode when SG is turned off, dark signal integrated in a photodiode (PD) is readout with low noise because of reset noise cancellation by correlated double sampling (CDS). However, during low conversion gain (LCG) mode when SG is turned on, bright signal integrated in the PD and CS is readout with larger noise because the reset noise is not cancelled by differential double sampling (DDS). The noise gap results in signal to noise ratio (SNR) drop at the conjunction point between HCG and LCG signals [2].

In this paper, a pixel circuit, which reduces the LCG reset noise without additional transistors in the pixel, is presented.

II. Overview of pixel circuit with Two Step Reset

Figure 1 shows a schematic of pixel array and peripheral circuits of the proposed LOFIC-CIS. A pixel is same as the conventional LOFIC-CIS which consists of a PD, a transfer gate (TG), a source follower transistor (SF), a select gate (SEL), a switching gate (SG), a sampling capacitor (CS) and a pixel reset gate (RST). In the conventional LOFIC-CIS, the reset noise hold in the CS after RST is turned off given by 643 μV_{rms} when CS = 10 fF, resulting in the SNR drop. A column reset circuit which consists of column reset gate (RST') and a column reset sampling capacitor (CS') are also located outside the pixel array. In this proposed LOFIC-CIS, RST and RST' are sequentially turned off to reduce the LCG reset noise, and this pixel circuit is named two step reset (TSR) pixel.

As shown in Fig. 2(a), readout operation timing is the same as the conventional LOFIC-CIS except for the LCG reset period. During the LCG reset period, both RST and RST' are turned on to reset floating diffusion node (FD) and the CS (t_1). The capacitance of the CS' is 1 pF that is 100× larger than that of CS so that temporal noise frozen on the RD node after RST' is turned-off is given by 64.3 μV_{rms} (t_2). The RST is then turned off (t_3). As long as RST is an ideal switch, the reset noise hold on CS will be 64.3 μV_{rms} because the RD node is high impedance. However, the charge injected to the CS flow to the RD node due to voltage drop difference between the CS and the RD node during the on-to-off transition time of RST gate as shown in Fig. 2(b), resulting in larger reset noise hold in the CS.

In order to reduce the reset noise in the CS, asymmetric structured RST is presented as shown in Fig. 3. With the tapered channel [3] shown in Fig. 3(a) and a half-buried channel shown in Fig. 3(b), the charge flow into CS is decreased. The voltage difference between the CS and the RD node when RST is turned off is thus decreased, so that the temporal noise caused by the current flow from the RD node to the CS is suppressed.

III. Simulation Results

Figure 4 shows transient noise SPICE simulation results of the TSR pixel. In order to confirm the noise caused by charge flow during the on-to-off transition time of RST, the fall time t_f is swept from 1 ps to 10 ns. The horizontal and vertical axes show t_f and the noise amount, respectively. When t_f is long, the noise of the TSR pixel is comparable to that of the conventional pixel, in which the theoretical reset noise is 643 μV_{rms}. However, when t_f is short, the noise of the TSR pixel is decreased from that of the conventional pixel, because the charge injected to the CS cannot fully flow into RD in the short time period. When t_f = 100 ps that is our target for RST clock driver, the reset noise is 375.2 μV_{rms} that is reduced by 28.3% compared to the conventional LOFIC-CIS.

The effect of the asymmetric structure RST is also feasibility studied with TCAD device simulations. Figure 5 shows the potential profile of the RST at the instance when RST is turned off. While potential shape in the channel of the normal structure RST is symmetric, the potential level in the CS is higher than that in the RD node. This difference of potential levels suggests that charge injected to small CS increases the potential level of the CS and charges flow to the RD node during the on-to-off transition. On the other hand, the potential level in the CS is close to that in the RD node with the asymmetric structure RST because the potential slope in the channel transfers the most of the charge to the large RD

node, and the increase of the CS potential is small. Since the potential difference between the CS and the RD node is small, it is expected that the current flow from the RD to the CS is suppressed and the reset noise is reduced.

IV. Evaluation result of a test chip

To verify the reset noise of the proposed TSR pixel, a test chip was fabricated with a 0.18 μm CMOS process. Figure 6 shows a photo of the test chip, in which $160(H) \times 4(V)$ pixels without PDs are implemented. Instead of the asymmetric RST with the tapered and half-buried channel structure, a simple asymmetric RST with additional contacts to the RD node shown in Fig. 7 is implemented due to fabrication process limitation. Since resistance value to the RD node is half of that to the CS, larger number of charges is expected to be injected to the RD node.

Figure 8 shows measurement setup of the test chip. A test chip (DUT), an 12-bit ADC and a FPGA are mounted on a PCB board. The power supply voltage V_{dd} and the pixel reset voltage V_R are provided by external power sources. The digitized DUT output with the ADC is transferred to the PC via USB and analyzed. Figure 9 shows the reset noise readout chain in the test chip, in which $v_{n,rst}$, $v_{n,SF}$, $v_{n,RO}$ and $v_{n,ADC}$ are respectively pixel reset noise, SF noise, readcout circuit noise and ADC noise. The measured noise with CDS operation is given by

$$v_{n,CDS}^2 = v_{n,SF}^2 + v_{n,RO}^2 + v_{n,ADC}^2, \tag{1}$$

where the pixel reset noise $v_{n,rst}$ is cancelled. On the other hand, the measured noise with DDS operation is given by

$$v_{n,DDS}^2 = 2v_{n,rst}^2 + v_{n,SF}^2 + v_{n,RO}^2 + v_{n,ADC}^2, \tag{2}$$

where the pixel reset noise is doubled due to uncorrelated reset noise. Therefore, the pixel reset noise is evaluated with measurement results at CDS operation and at DDS operation, which is given by

$$v_{n,rst} = \sqrt{\frac{v_{n,DDS}^2 - v_{n,CDS}^2}{2}}. \tag{3}$$

The pixel reset noise is summarized in Table I. In the case of the conventional pixel, the measured reset noise is 1028 μV_{rms}, even though it is larger than theretical value of 643 μV_{rms}. In the case of the TSR pixel with a normal structure RST, the measured reset noise is 734 μV_{rms} that is reduced by 28.6% compared to the conventional pixel. The noise reduction ratio is comparable to the simulation result. In the case of the TSR pixel with the asymmetric structure RST, the measured reset noise is further reduced to 607 μV_{rms}, that is 41.0% lower than that of the conventional pixel. This result suggests that the reset noise is reduced by suppressing the injection charge flow from the CS to the RD node. Therefore, it is expected that the TSR pixel with asymmetric RST composed the tapered and half-buried channel can reduce the reset noise further because the number of charge injected to the CS will be lower than that of the simple asymmetric RST implemented.

V. Summary and future work

In order to realize a LOFIC CIS with small SNR drop at the conjunction point between HCG and LCG signals, the TSR pixel circuit, which reduces pixel reset noise without additional transistors in a pixel has been proposed. The LCG reset is conducted by two-step with the additional reset transistor RST' and the reset sampling capacitor CS' located outside the array. The temporal noise on the reset drain RD is frozen with large CS', and the reset noise on the CS is suppressed as long as the charge injected to the CS does not flow to the RD node during the on-to-off transition of RST.

SPICE noise simulation results show that the pixel reset noise is suppressed by 28.3% at $t_f = 100$ ps. In order to further reduce the noise, the asymmetric RST with tapered and half-buried channel is proposed. According to TCAD simulation result, it is confirmed that the potential difference between the CS and the RD node is reduced. It is expected that results in reduction of injection charge flow form the CS to the RD node.

The test chip of the TSR pixel with a simple asymmetric RST is fabricated with 0.18μm CMOS process. Evaluation results show that the pixel reset noise is reduced by 41.0%.

As future work, we will fabricate a prototype chip of the TSR LOFIC-CIS with the asymmetric RST with tapered and half-buried channel.

VI. Acknowledgments

References

[1] S. Sugawa, N. Akahane, S. Adachi, K. Mori, T. Ishiuchi, and K. Mizobuchi, "A 100 db dynamic range cmos image sensor using a lateral overflow integration capacitor," in *ISSCC. 2005 IEEE International Digest of Technical Papers. Solid-State Circuits Conference, 2005.* IEEE, 2005, pp. 352–603.
[2] N. Akahane, S. Adachi, K. Mizobuchi, and S. Sugawa, "Optimum design of conversion gain and full well capacity in cmos image sensor with lateral overflow integration capacitor," *IEEE transactions on electron devices*, vol. 56, no. 11, pp. 2429–2435, 2009.
[3] J. Ma and E. R. Fossum, "A pump-gate jot device with high conversion gain for a quanta image sensor," *IEEE Journal of the Electron Devices Society*, vol. 3, no. 2, pp. 73–77, 2015.
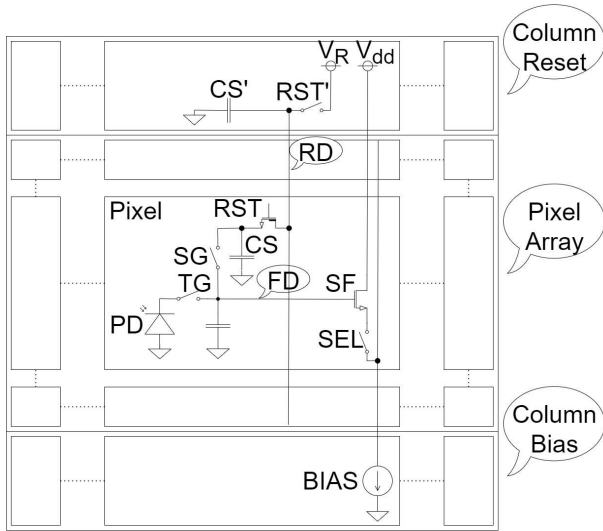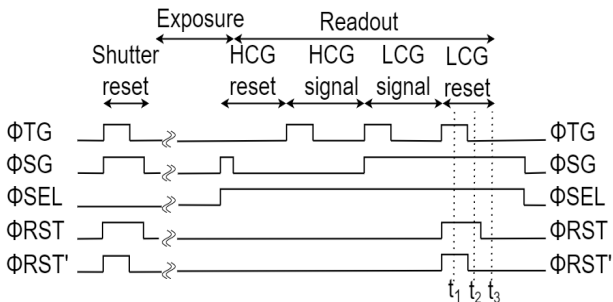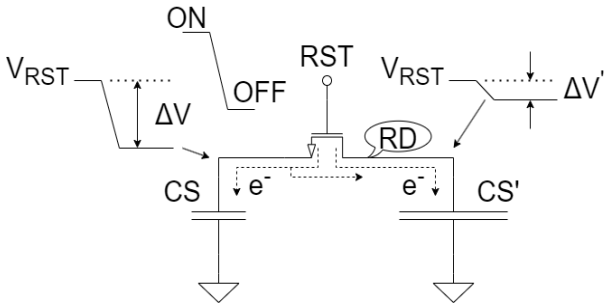
Fig. 1. Schematic diagram of the proposed LOFIC-CIS
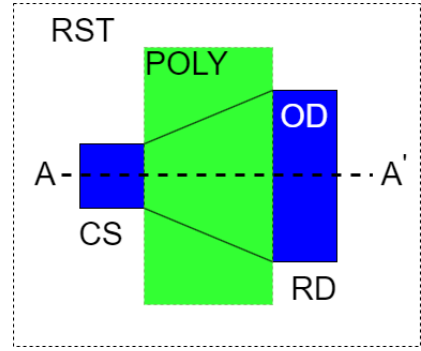

(a) Timing diagram of the proposed LOFIC-CIS

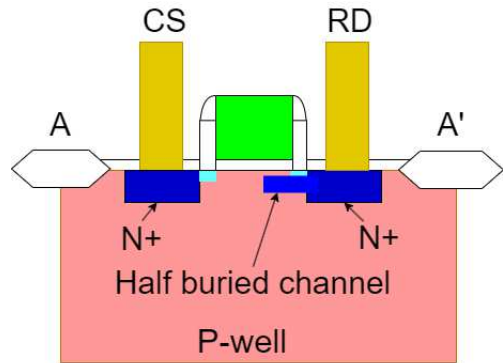(b) Injected charge to the CS flows to the RD node due to voltage drop difference

Fig. 2. Timing diagram and injected charge flow

(a) Top view of tapered channel

(b) A-A' cross-sectional view of half-buried channel

Fig. 3. Asymmetric structured RST

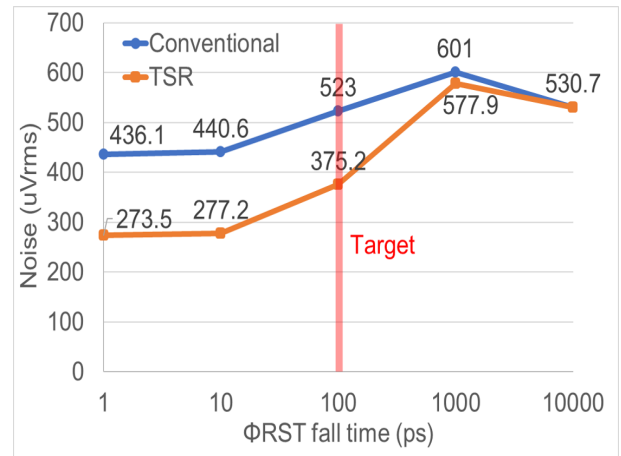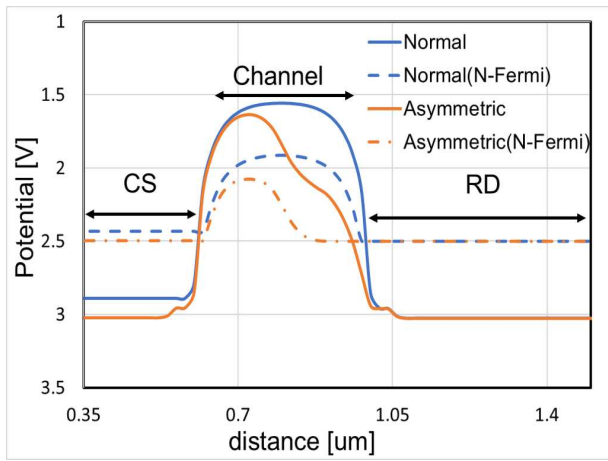Fig. 4. Transient noise SPICE simulation results

TABLE I
EVALUATION RESULT OF PIXEL RESET NOIES.

	Conventional	TSR (normal)	TSR (asymmetric)
Noise	1028 μV_{rms}	734 μV_{rms}	607 μV_{rms}

Fig. 5. TCAD device simulation result of asymmetric strcutured RST
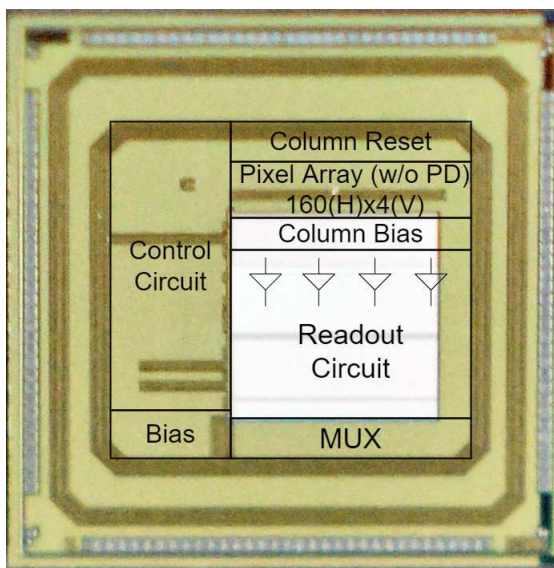
Fig. 6. A photo of the test chip

Fig. 8. Measurement setup of the test chip

Fig. 7. Alternative simple asymmetric RST with double contacts to the RD node
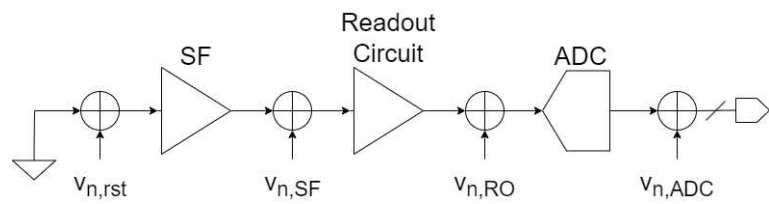
Fig. 9. Reset noise readout chain of the test chip measurement

High Precision Direct ToF Ranging using CMOS SPAD and Ultra-Short Pulsed Laser

Tsai-Hao Hsu*, Chun-Hsien Liu, Tzu-Hsien Sang, and Sheng-Di Lin

Institute of Electronics, National Yang Ming Chiao Tung University, 1001 University Road, Hsinchu 30010, Taiwan.
Corresponding Author: h124129859.s@gmail.com, telephone: +886-3-5712121#54240

Record single-shot precision of <20 μm for direct-ToF ranging in sub-second integration time has been achieved. By varying the excess bias of SPAD, we found that the FWHM of collected histogram plays a deterministic role on the ranging precision. The effect of target reflectivity, laser repetition rate, and integration time on precision has been unified as a single parameter named effective integration time, T_{eff}. A simple relation between T_{eff} and precision has also been proposed and verified.

I. Introduction

Light-detection and ranging (LiDAR) system plays a key role in various application like auto-driving and satellite ranging, where CMOS single-photon avalanche diodes (SPADs) can be one of the best detectors in the receiver end because of its single-photon sensitivity, excellent timing resolution and easy fabrication in CMOS process. Fast measurement with high ranging precision could trigger various applications [1,2]. In this paper, we investigates the precision of SPAD-based time-of-flight (ToF) ranger as the race on high-precision distance measurement has been ongoing for decades. In terms of cost and system complexity, direct ToF (d-ToF) is highly competitive among various ranging setups and the milli-meter level precision at 50 m distance have been achieved [4-8]. In this work, we demonstrate a record precision of 15.6 μm in half second integration time (T_{int}) with a d-ToF ranger using a CMOS SPAD chip, a 70-ps short-pulse 905-nm laser, and a 10-ps time-bin resolution time-to-digital converter in a TCSPC card.

II. Experiment setup and result

A single SPAD in a 64x128 array chip fabricated with 180-nm BCD process without any customization was used for this ranging experiment [3]. Figure 1 show the chip layout. The chip size including 64x129 SPADs and their quenching, reset, and readout circuits, is about 4.1x5.0 mm². The column and row selectors allow us to activate any single SPAD. The respective active size and breakdown voltage are 14 μm and ~ 49.5 V. The dark-count rate, photo-detection probability (PDP) at 905 nm, and deadtime at 6-V excess bias (Vex) are 1 kHz, 9%, and 10 ns, respectively. Thanks to the high sensitivity of SPAD and to avoid pile-up effect in timing histogram, the focal lens in front of the receiver was removed. Figure 2 illustrated our ranging setup including the TCSPC with time-tag time-resolved (TTTR) function to record every count in a long T_{int} (T_L, 100 or 200 seconds here) for subsequent analysis. Figure 3a shows the normalized T_L histograms using 90% target reflectivity (Rt) and 1-MHz laser at the excess bias of 1.7 – 6 V, which resulted the histogram full-width half maximum (FWHM) in the range of ~1300 – 160 ps on the right axis in Figure 3b, together with the bias-dependent PDP and return probability on the left y-axis, denoted as RP, which is defined as the average return counts per fired laser pulse. The PDP in the bias range is about 2%-5%. The RP in the range of

8% - 17% indicates a negligible pile-up effect in the timing histogram, which is particularly important to study the effect of histogram FWHM on the ranging precision.
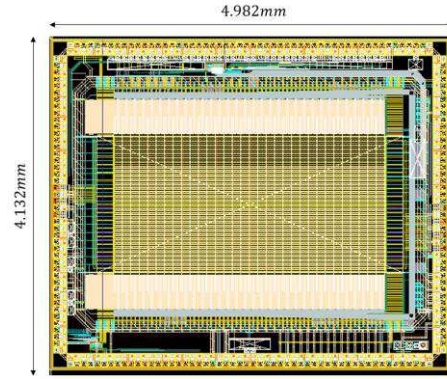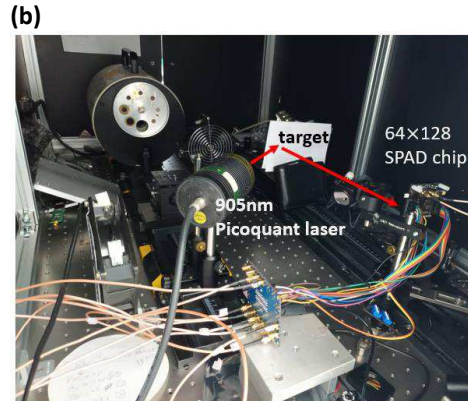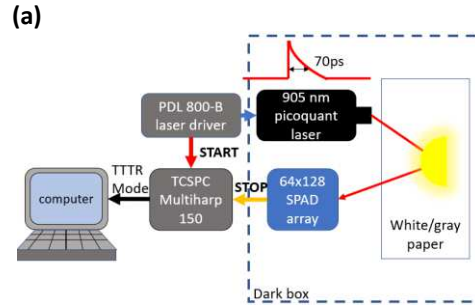

Fig. 1 Layout of the 64X128 SPAD chip.

(a)

(b)

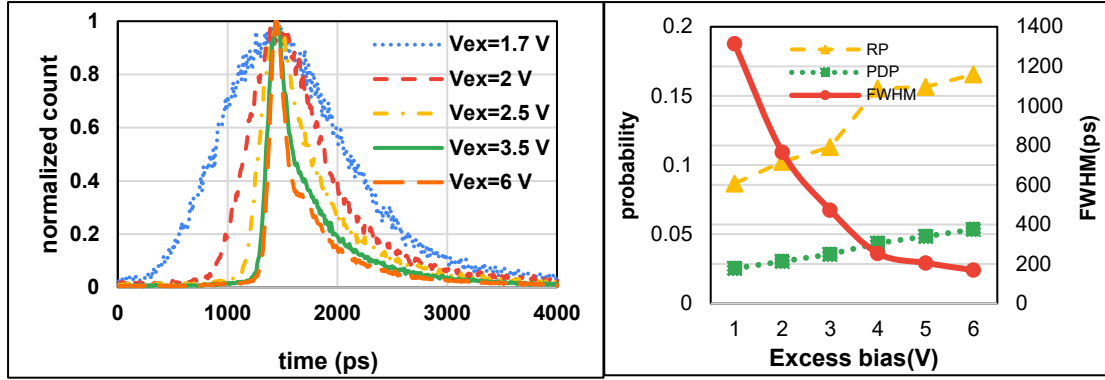**Fig. 2 (a) Block diagram of experiment setup, (b) ranging setup photo.**

Fig. 3 (a) the normalized histogram for 1 second integration time in various Vex, (b) the RP、PDP and FWHM of histogram in different excess bias.

Using center-of-mass (CM) method for peak detection, we have obtained the measured distance with the TTTR data for all integration times. Figure 4 shows the standard deviations from >100 measurements, defined as the ranging precision (σ_{CM}), as a function of Tint. The precision σ_{CM} improved dramatically with the increasing Tint. Clearly, the best precision of 15.6 μm was achieved successfully with 6 V Vex, 10 MHz laser, Rt = 90%, and Tint = 0.4 s. The inset on Figure 4 demonstrates that, as expected, the probability density functions of the measured distances follow respective Gaussian distributions very well and the increasing Tint gave the smaller σ_{CM}.
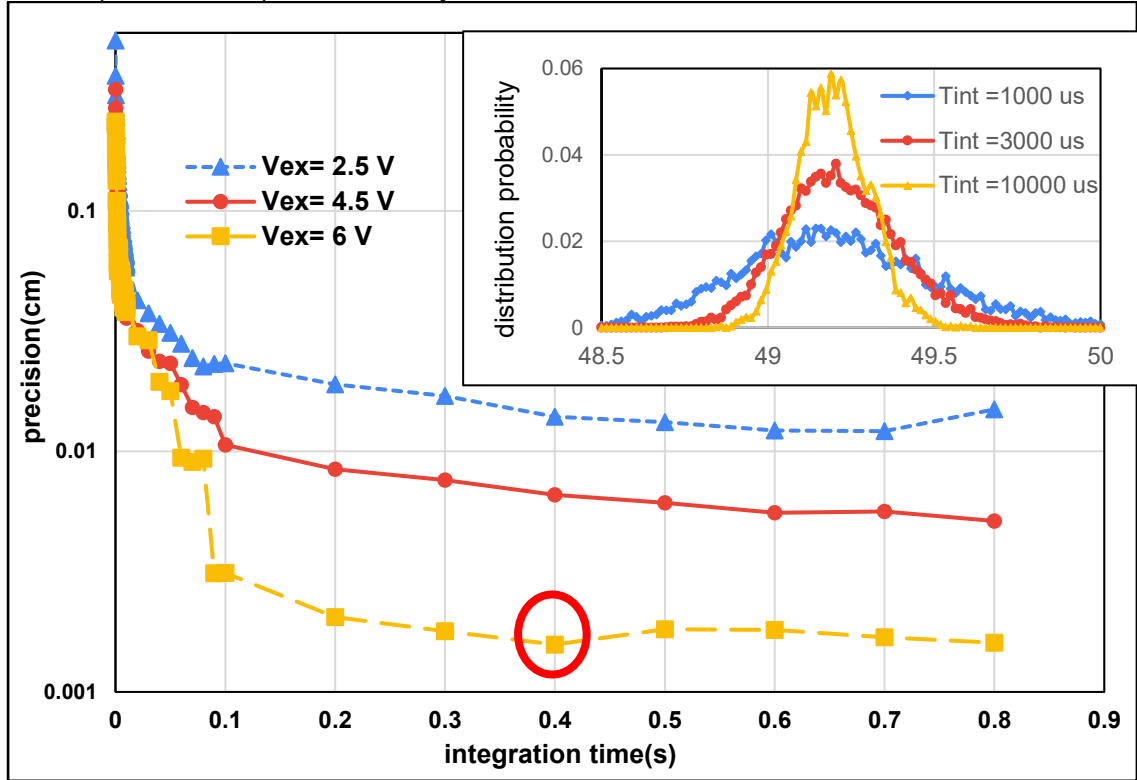

Fig. 4 Precision vs. integration time at various Vex. Inset: measurement distribution at different integration times.

III. Theoretical model for ranging result

Figure 5a plots the precisions σ_{CM} as a function of FWHM of T_L histograms obtained with three T_{int}. It can be seen that, irrespective of varied T_{int}, σ_{CM} decreases with the decreasing FWHM. To clarify the key factors affecting ranging precision, we introduce the effective integration time, denoted as T_{eff}, defined as,

$$T_{eff} = T_{int} \times RP \times f_L, \qquad (1)$$

where f_L is the repetition rate of laser. In fact, T_{eff} is the valid laser counts triggered by the returned photons from target scattering so its physical unit is counts. In this way, we can exclude the effect of the target reflectivity, the SPAD PDP, and the laser repetition rate. Fig. 5b shows a plot similar to that in Fig. 5a but at three T_{eff} instead. The obtained trend is very similar, too, indicating that the FHWM of timing histogram is the dominate factor in ranging precision.
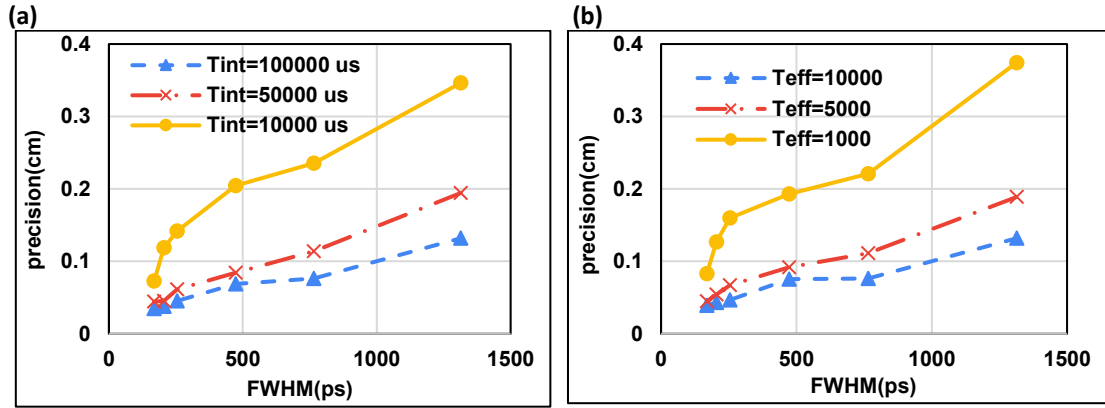
Fig 5. FWHM vs. precision.

Figure 6 clearly exhibits the dominant role of histogram FWHM by plotting the precision at three Vex (or 3 histogram FWHMs) as a function of T_{eff} for a few different ranging conditions, including $R_t = 18\%$ or 90% and $f_L = 1$ or 10 MHz. Obviously, at the same Vex (or the same histogram FWHM), the relation between the measured precision and T_{eff} is the same as all data points fall on the same line. Interestingly, the log-log plot reveals an interesting dependence,

$$\sigma_{CM} \propto T_{eff}^{\frac{-1}{2}}, \tag{2}$$

After taking the fluctuation of counts in each time bin and calculating the error propagation to our CM method [1], we can approximate the precision σ_{CM} as,

$$\sigma_{CM} \cong \sqrt{\sum_{n=-\frac{N}{2}}^{n=\frac{N}{2}} (\frac{n}{c_t})^2 c_n} \propto c_t^{-1/2}, \tag{3}$$

where explaining the slope = -0.5 in log-log plot in Fig. 6..

Figure 7 summarizes the d-ToF precision as a function of ranging precision [2]. Clearly, our precision is the best among reported works and it could be of high potential for future ranging applications.
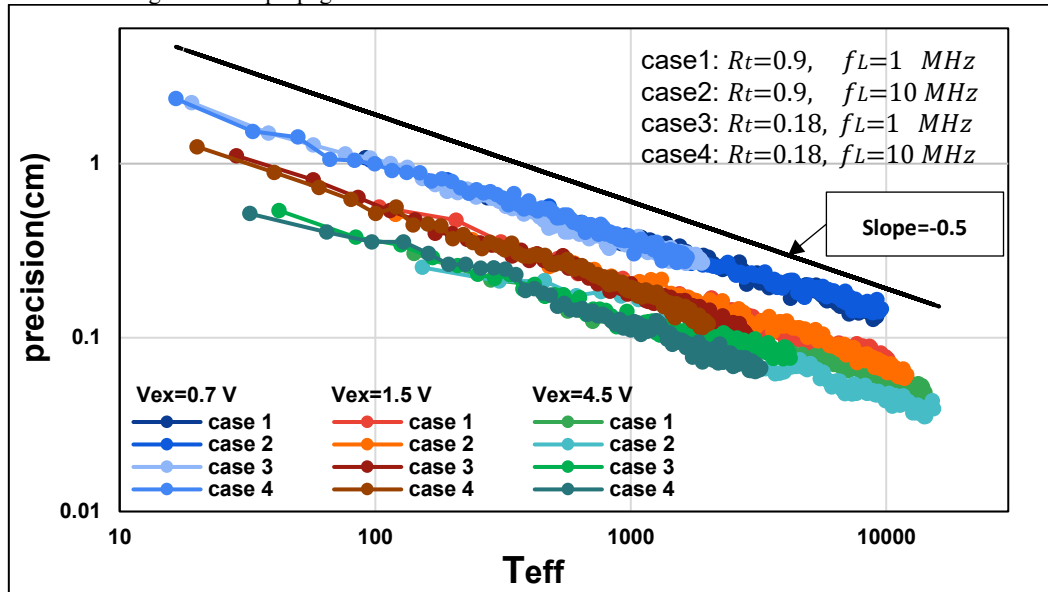

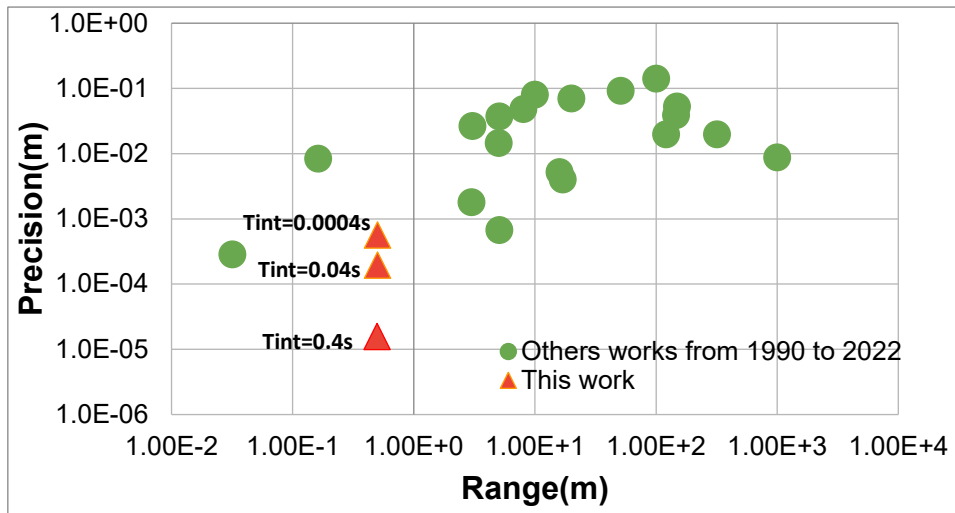**Fig. 6 Precision vs. T_{eff} at three Vex and in four cases in log-log scale**

Fig. 7 Precision vs. operating range since 1990 to 2022.

IV.　Conclusion

In this work, with 0.4-s Tint、70-ps laser FWHM and 6-V Vex, we have obtained 15.6 μm precision in d-TOF SPAD LiDAR at about 50-cm distance with low background condition. A theoretical model to estimate the relationship between the number of detected photon and precision has been proposed to explain our experimental result.

References

[1]　N. Li, C. P. Ho, J. Xue , L. W. Lim, G. Chen, Y. H. Fu, and L. Y. T. Lee "A Progress Review on Solid-State LiDAR and Nanophotonics-Based LiDAR Sensors." Laser & Photonics Reviews, vol. 16, August 2022.

[2]　B. Behroozpour, P. A. M. Sandborn, M. C. Wu, and B. E. Boser, "Lidar System Architectures and Circuits," in *IEEE Communications Magazine*, vol. 55, no. 10, pp. 135-142, October. 2017.

[3]　P. Chen, C. Liu, A. Hsiao, Y. Tsou, Y. Fang, L. Ko, H. Tsai, C. Tsai, T. Sang, G. Lin, J. Guo, B. Hsiao, and S. Lin, "Minimum ranging time for a LiDAR module using CMOS single-photon avalanche diodes," in Conference on Lasers and Electro-Optics, Technical Digest Series (Optica Publishing Group, 2022), paper JW3A.13, May 2022.

[4]　C. Niclass, A. Rochas, P. . -A. Besse and E. Charbon, "Design and characterization of a CMOS 3-D image sensor based on single photon avalanche diodes," in IEEE Journal of Solid-State Circuits, vol. 40, no. 9, pp. 1847-1854, Sept. 2005

[5]　S. Kawahito et al., "A CMOS Time-of-Flight Range Image Sensor with Gates-on-Field-Oxide Structure," IEEE Sensors J., vol. 7, no. 12, pp. 1578–86, Dec. 2007.

[6]　Z. Chao, S. Lindner, I. M. Antolovic, M. Wolf, and E. Charbon, "A CMOS SPAD Imager with Collision Detection and 128 Dynamically Reallocating TDCs for Single-Photon Counting and 3D Time-of-Flight Imaging," Sensors, vol. 18, Nov. 2018.

[7]　C. Zhang, S. Lindner, I. M. Antolović, J. Mata Pavia, M. Wolf and E. Charbon, "A 30-frames/s, 252×144　SPAD Flash LiDAR With 1728 Dual-Clock 48.8-ps TDCs, and Pixel-Wise Integrated Histogramming," in IEEE Journal of Solid-State Circuits, vol. 54, no. 4, pp. 1137-1151, April 2019.

[8]　T. Paweł, D. Z. Wziątek, S. Dalyot, T. Boski, and F. P. L.Filho, "A High-Precision LiDAR-Based Method for Surveying and Classifying Coastal Notches," ISPRS International Journal of Geo-Information, vol. 7, no. 8, pp. 295, July 2018.

High-speed, super-resolution 3D imaging using a SPAD dToF sensor

Germán Mora-Martín [1], Jonathan Leach [2], Robert K. Henderson [1], Istvan Gyongy [1]

[1]The University of Edinburgh, Institute for Integrated Micro and Nano Systems, Edinburgh, U.K.
[2]Heriot-Watt University, Institute of Photonics and Quantum Sciences, Edinburgh, U.K.
Istvan.Gyongy@ed.ac.uk Tel: +44 131 651 7054

Abstract—High-speed 3D time-of-flight (ToF) imaging has the potential to offer improved situational awareness in robotics and automotive applications as well as assisting photogrammetry-based high-speed scientific imaging such as material testing. This paper uses a CMOS SPAD dToF sensor for depth and intensity imaging at up to 10kFPS. Depth maps are upscaled from a resolution of 64×32 to 256×128 using a recently proposed video super-resolution techniques tailored to SPADs. We also present preliminary results from the application of the sensor to human activity recognition (HAR).

I. INTRODUCTION

The use of SPAD-based 3D depth sensors has become widespread in the last few years, with the sensors finding applications in smartphones, robotics, and even home appliances [1]. SPADs have also become a key technology in LIDAR for autonomous systems [2]. By integrating SPAD arrays with processing logic, solid-state, all-digital receivers can be implemented that provide accurate depth maps even in high ambient conditions. However, array sizes tend to be limited, leading to a relatively low angular resolution when imaging in a flash modality. Instead of using flood illumination, some SPAD modules project a dot array (using a diffractive optical element [3]) which increases the SNR in the spots and thus the range but again results in sparse spatial sampling. There is therefore an interest in using post-processing to improve the lateral resolution of depth maps, as well as to provide scene interpretation, especially for long-range targets subject to significant pixelation.

II. SENSOR ARCHITECTURE

We used a high-speed SPAD dToF sensor in our study [4], capable of running at frame rates in the 10 kFPS range (>100 kFPS for on-chip depth computation). The sensor, implemented in STMicroelectronics' 40nm technology, comprises 64×32 pixels, each pixel consisting of a 4×4 array of SPADs and processing logic. A time-gated, multi-event histogramming TDC is integrated into each pixel, generating an 8-bin histogram with a resolution down to ~250 ps [4]. The time gate functionality enables the histogram to be shifted in time to extend the range of the sensor. Three main mechanisms are available for setting the time gate positions of individual pixels: (1) internal control that automatically tracks the signal peak via in-pixel background estimation and peak detection, (2) internal control that continually cycles across up to 128-time gate positions, and (3) external control, potentially based on guidance from an additional sensor [5] (such as a stereo vision system). On-chip (column parallel) depth computation and selective readout options are available to provide further data compression. In addition to time-resolved imaging, the sensor offers a 128×128 photon counting (intensity) imaging modality. Figure 1 shows a portable camera setup consisting of a custom PCB (housing the SPAD and an FPGA module), a compact 850 nm VCSEL illumination module, and a laser range finder for reference depth measurements. The camera is connected to a laptop which controls and powers the camera; a Matlab software interface provides real-time visualisation of the captured data. For post-processing, we used a desktop computer (HP EliteDesk 800 G5 TWR) with an RTX2070 GPU.

III. HIGH-SPEED, SUPER-RESOLVED DATA

Figure 2 depicts examples captured at 200 FPS in the tracking modality (mode 1 above) of the sensor. The scene is of two people, one running and the other waving, in an open space with objects scattered around. The figure shows depth data obtained by applying centre-of-mass peak extraction on the histogram frame (panel a) as well as the upscaled version (from 64×32 to 256×128) of this data following neural network super-resolution processing (b) [6], which is seen to lead to an improvement in the profiles of the people. Unlike commonly used, intensity-guided approaches [7], the upscaling is based entirely on the depth data here. Processing speeds above 30 FPS are achieved. Figure 3 shows data acquired indoors at 10 kFPS of a balloon being burst. Three sequences are given: intensity frames (3a), depth frames (3b), and super-resolved depth data (3c). The rupture of the balloon is captured in high temporal detail, demonstrating the potential of SPAD cameras in specialised high-speed imaging applications [8], especially where high sensitivity is required, as offered by state-of-the-art SPADs [9]

Figure 1. a) Picture of the camera setup. A 25 mm/f1.4 objective is used in front of the sensor, giving a 20×5 degree field-of-view (FOV), together with a 10 nm ambient filter. Illumination is provided by a compact 850 nm VCSEL source with 10 ns pulse width and 60 W peak optical power triggered at 1.2 MHz. b) The FOV of the camera.
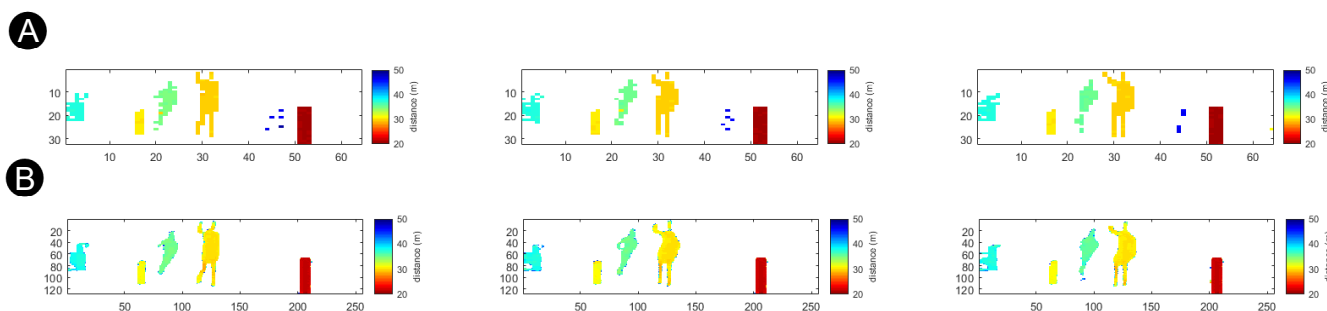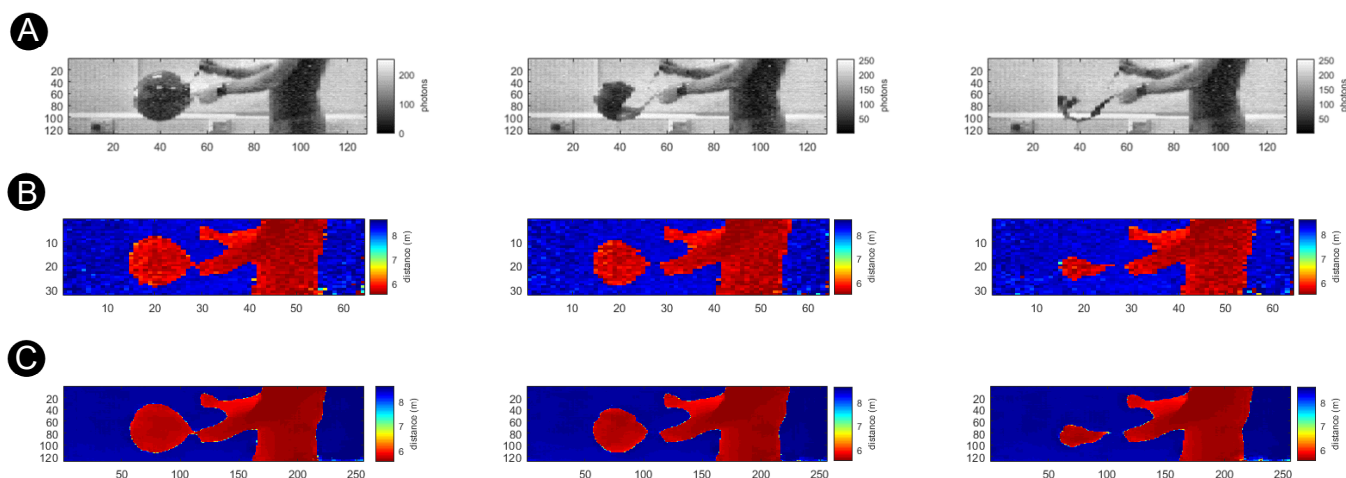

Figure 2. Selected frames from data captured in tracking mode at 200 FPS with 8 ns bin size and 16 time gate positions (giving 81.6m of unambiguous range) a) depth maps obtained by centre-of-mass processing of histogram frames b) corresponding super-resolution depth maps. Only pixels which are detecting surfaces in the 20-50 m range are plotted in panel (a).

64×32 depth data derived from histogram frames c) the data in panel b upscaled to 128×256 after super-resolution processing. Note that the photon counting, and depth sequences were captured separately.

IV. HUMAN ACTIVITY RECOGNITION

Human activity recognition (HAR) has gained importance in computer vision due to its applications in video surveillance, health care services, human-computer interaction, and autonomous driving [10]. Using just depth information for HAR has become a popular research topic due to the preservation of privacy and fast speed compared with other methods (e.g., using 3D skeletons) [11]. Furthermore, depth-based detection has the potential to work even when there is no colour contrast between the person and the background (for example when camouflage is used [12]). However, one of the key challenges is to overcome the low transverse resolution of depth maps when imaging from a distance.

Recurrent neural networks (RNNs) are a very useful type of network for sequential data. In particular, convolutional long-short-term memory (Conv-LSTM) layers are key to learning spatio-temporal features from data [13]. In [11], an RNN based on conv-LSTM layers is used to perform HAR on high-resolution, indoor, and short-range depth data. In this work, we use a similar network to perform HAR on data from a SPAD dToF sensor involving longer-range, outdoor sequences, where SNR is typically lower, and objects can become heavily pixelated.

The method is designed to perform HAR on sequences of any length. First, a 64×32 depth sequence is captured and passed through a U-net-like network to localise people [14]. Next, the depth sequence is cropped spatially around each person in frames of 16×32 pixels and then resized to 32×32 pixels. Finally, the cropped sequence is analysed by the RNN network and outputs an activity from the following set: remaining idle, walking, running, crouching down, standing up, waving, or jumping. Figure 4 shows a diagram summarising the steps involved in this approach to perform HAR.

Unreal Engine [15] was used to generate a large and diverse training dataset, shared for both human localisation and HAR networks. Ground truth information for a variety of sensors can be extracted from virtual environments. In this work, depth, intensity, and segmentation frames of size 512×128 are recorded and used as inputs in an optical model to simulate data from a SPAD dToF sensor. To match the sensor architecture described here, the model assumes 4×4 SPAD macropixels, a pixel resolution of 64×32 (with an aspect ratio of 4:1), and in-pixel histogramming.

Figure 4 shows the confusion matrix of all activities considered here, indicating % of samples predicted in a given class in the test dataset (data unseen by the model). The overall accuracy for the test dataset is 91.5%. Activities corresponding to standing up, walking, running, jumping, and waving are detected with a recall higher than 90% while crouching down, and remaining idle has lower recall values (though a precision of >93%). False positives can occur due to similarities between two actions (such as crouching down slightly before jumping) or failure to localise the person accurately (in some cases due to distracting features in the background). The network is able to perform HAR from a sequence with a latency of 150 ms.

Activity sequences were captured using the SPAD dToF sensor at 50 FPS to generate a test dataset from real data. Figure 6 compares a sequence of a person walking captured by the real dToF sensor (Fig. 6a) with corresponding synthetic SPAD data (Fig. 6b). The visual similarity between the two sequences appears to justify the use of synthetic SPAD data for training. Indeed, preliminary results suggest similar performance on the real dataset to the results on the synthetic dataset, HAR predicting activities such as running, walking, and standing up with high sensitivity, whilst crouching down and remaining idle have reduced recall values (but high precision).

Prediction

Truth \ Prediction	Crouching	Idle	Waving	Running	Walking	Jumping	Standing
Crouching	84.8	0.8	0	0	0	9.7	4.7
Idle	0.5	83	1.1	0.6	8.5	4	2.3
Waving	1.7	3.9	90	0	2.2	0.5	1.7
Running	0	0	0	99.4	0.6	0	0
Walking	0	0	0	6.5	92.5	0	1
Jumping	1.6	0.5	0	0	0	92	5.9
Standing	0	0	0	0	0	0	100

Figure 4: Confusion matrix of activities representing % of samples predicted in each class. Example: 1.6% of jumping data is confused with crouching down. Each class has approximately 200 samples.

V. CONCLUSIONS

We demonstrated the application of a dToF SPAD sensor in high-speed imaging and showcased the use of deep learning models, trained on synthetic SPAD data, to overcome the limited transverse resolution and provide upscaled depth maps or human activity recognition (HAR). Future work will attempt to improve segmentation and extend the method to multiple people within the field of view.

REFERENCES

[1] https://www.st.com/content/st_com/en/about/media-center/press-item.html/t4210.html *Last visited 06/04/2023*

[2] Li et al., *Federated learning: Challenges, Methods, and Future Directions,* IEEE Signal Process. Mag, 37(4), 2020

[3] *Breaking Down iPad Pro 11's LiDAR Scanner*, EE Times. Available online: https://www.eetimes.com/breaking-down-ipad-pro-11s-lidar-scanner

[4] Gyongy et al., *A direct time-of-flight image sensor with in-pixel surface detection and dynamic vision,* JSQTE 2023

[5] Taneski et al., *Guided Flash Lidar: A Laser Power Efficient Approach for Long-Range Lidar,* IISW 2023

[6] Mora Martín et al., *Video super-resolution for single-photon LIDAR,* Optics Express 31, 7060-7012 2023

[7] Ferstl et al., *Image guided depth upsampling using anisotropic total generalized variation,* ICCV 2013

[8] Etoh et al., *Needs, requirements, and new proposals for ultra-high-speed video cameras in Japan*, ICHSPP 1994

[9] Shimada et al., *A SPAD depth sensor robust against ambient light: the importance of pixel scaling and demonstration of a 2.5 μm pixel with 21.8% PDE at 940 nm,* IEDM 2022

[10] Song et al., *Pattern recognition, ICPR international workshops and challenges,* ICPR 2021

[11] Sánchez-Caballero et al., *Real-time human action recognition using raw depth video-based recurrent neural networks,* Multimedia Tools and Applications 2022

[12] Tachella et al., *Real-time 3D reconstruction from single-photon LIDAR data using a plug-and-play point cloud denoisers,* Nature Communications 2019

[13] Xingjian et al., *Convolutional LSTM network: a machine learning approach for precipitation nowcasting,* Advances in neural information 2015

[14] Ronneberger et al., *U-Net: convolutional networks for biomedical image segmentation,* MICCAI 2015

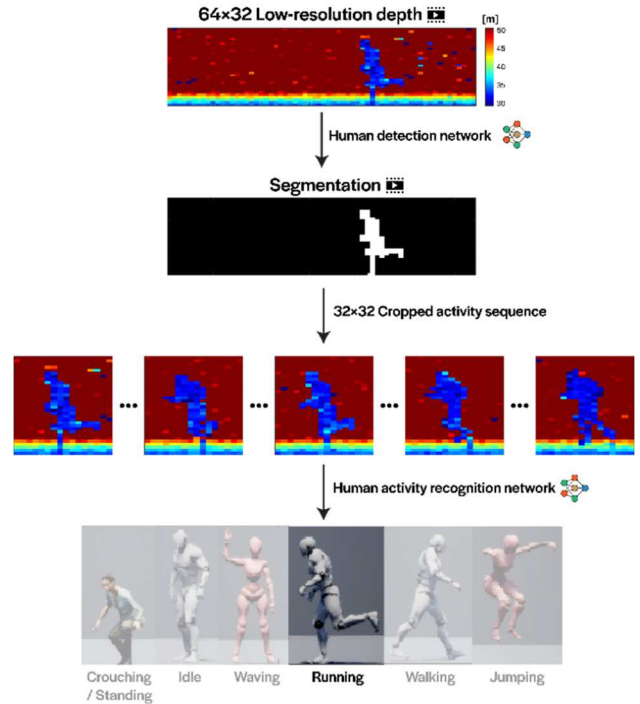[15] Epic Games, *Unreal Engine*

Figure 5. Human activity recognition workflow diagram. A low-resolution depth sequence is captured and segmented via a human segmentation network (similar to U-net). Based on the localisation of the human, the sequence is cropped accordingly. The cropped sequence is passed through a second network evaluating the activity performed (e.g. running).
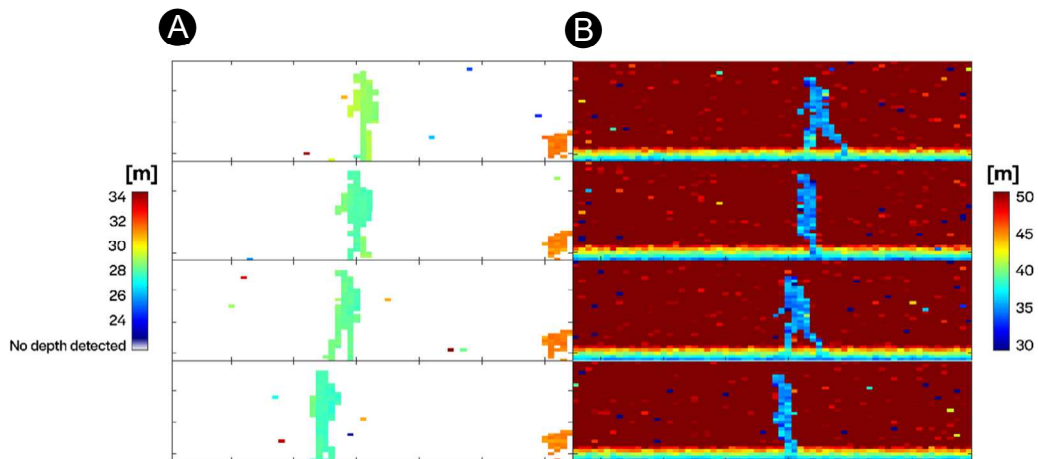

Figure 6. Comparison of selected frames from a sequence of a person walking captured with a) a real SPAD camera (50 FPS, mean signal-to-background photons ratio (SBR) of 0.17 and 716 average signal photons for the person in the first frame) b) a virtual SPAD camera (model-generated data, mean SBR 0.19 and 42 average photons).

Self-Powered Ambient Light Sensor Using Energy Harvesting Pixels and Zero Power Communication

B. Sarachi[1,2], A. Cook[2,3], J. M. Raynor[2], I. Todorova[2,4], S. Ball[2,4], F. Kaklin[1,2], J. MacDougall[2,3], S. M. Aparicio[1,2], R. K. Henderson[1]

[1]School of Engineering, Institute for Integrated Micro and Nano Systems, The University of Edinburgh, s1803356@ed.ac.uk

[2]STMicroelectronics Imaging Division, Edinburgh, UK

[3]School of Engineering & Physical Sciences, Heriot-Watt University, Edinburgh

[4]School of Engineering, University of Glasgow, Glasgow

INTRODUCTION

As CMOS technology improves, more efficient devices can be created achieving functionality with ever decreasing power budgets. This has moved to the point where some devices are so low power that they are able to harvest what they require from their surroundings. These energy harvesting sensors can be implemented in a wider system providing functionality with very little power cost. We present one such system, an energy harvesting (EH) ambient light sensor (ALS). This device is a 3D stacked chip where the top tier consists of an array of PV cells to harvest energy from the ambient light and a photodiode sensing cell to provide the ALS measurement. Having both power generation and sensing functionality contained within the same array is possible with deep trench isolation (DTI). The full thickness trench isolation allows each PV cell to be isolated from one another and then connected in series allowing for high voltage generation without the use of charge pumps [1]. By using a back side illuminated process (BSI) both the quantum efficiency (QE) and fill factor are improved. In this case, 3 PV cells are connected in series providing 1.2V to power the device forming the basis for a self-powered ambient light sensor. This sort of sensor has uses in many applications in consumer electronics that already use an ALS and have limited battery power such as wearables and small IoT devices.

POWER BUDGET DERIVATION

Since the ALS has no other voltage source, its power budget is defined by the amount of energy the PV cells can harvest at minimum irradiance. The amount of energy available from ambient light is dependent on the spectra of its source. When defining the power budget for the device, a worst-case of fluorescent light was used. This is because fluorescent light has lower power density compared to other light sources like solar and LED. By integrating the irradiance over the range of wavelengths within the fluorescent light spectra [2], the total power density was found to be 40.4W/m^2. This means that 100 lux of fluorescent light produces 0.297 W/m. The efficiency of the harvesting cells is 10% [3] at low light levels ($\approx$100 lux). The top tier area is defined as 2mm by 2mm with the sensing PV being 10um x 10um.

$$P = F * A$$

P is the power generated, F is the power density of the light source, and A is the cell area. Therefore, the power budget for the chip is 120nW for 100lux of fluorescent light.

ANALOGUE DESIGN

As shown in figure 1, the analogue circuitry on chip is responsible for taking the photodetector measurement of the ambient light and digitising it so it can be serially communicated off-chip. The photodetector itself is forward biased which allows for a higher sensing dynamic range. In this configuration, a voltage signal which is logarithmically dependent on the light level is output and can directly be fed directly into the ADC. The approach to the analogue design on chip was to reduce the power consumption as much as possible which often meant omitting parts that were not strictly necessary to the functionality of the ALS. By having a forward biased photodetector directly connected into the ADC there was no need for a buffer or sample and hold circuitry. Another measure taken to reduce power consumption was to lower the frequency of the clock generated on chip which coupled with using longer transistors meant a reduction in leakage currents.

The 3 PVs in series that make up the supply voltage each produce 0.413V at minimum irradiance of 0.33W/m^2. This figure increases to 0.627V at maximum irradiance of 250W/m^2. The voltage reference generator used was not a bandgap reference circuit as the minimum supply voltage was too low at 1.2V. Therefore, the current and voltage references are derived from the power supply voltage. This means the voltage reference generator output is dependent on light level but the dynamic range of the light incident on the device is high enough that this dependency is acceptable.
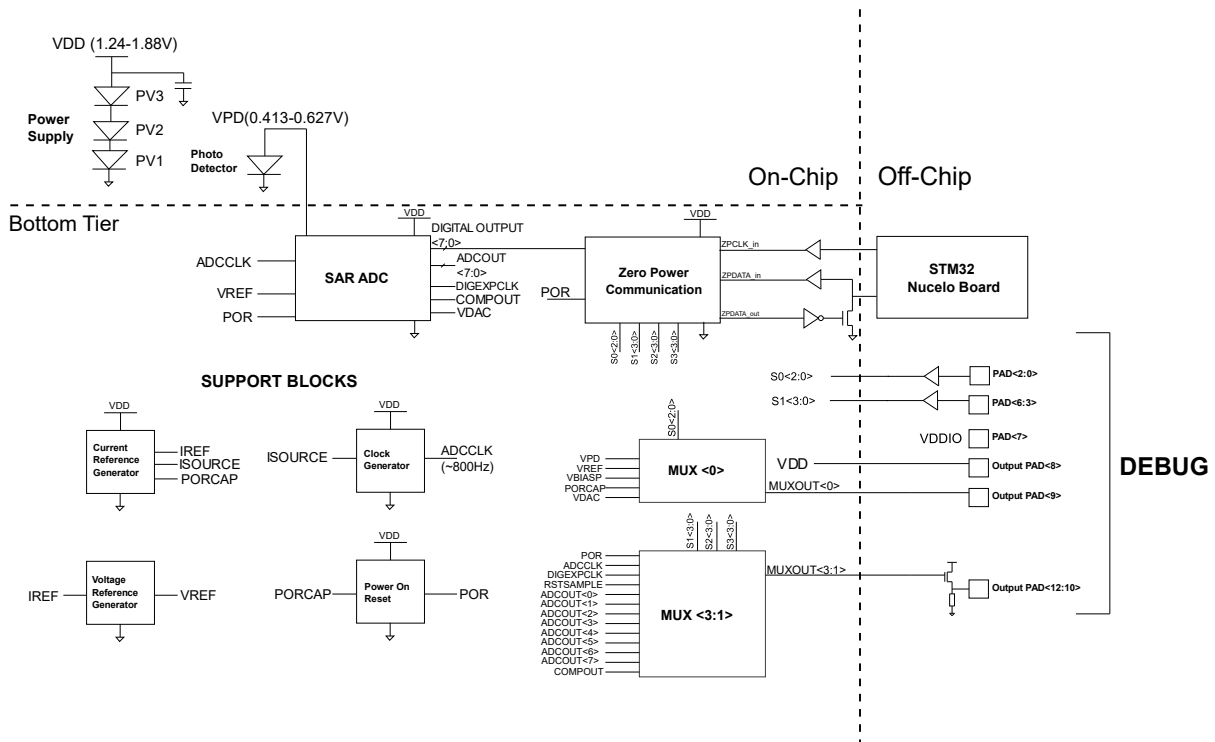
Figure 1: Chip block diagram. Debug pads included for testing.

The ADC used in this design is a successive approximation register (SAR) ADC using a dynamic latch comparator. This was implemented because the comparator has low static power consumption as it only consumes power during comparisons. When idle, the comparator consumes negligible power. To conserve power, the ADC maximum sampling frequency is 36Hz.

ZERO POWER COMMUNICATION

The purpose of the digital circuitry in the design is to take the parallel digital output of the ADC, serialise it and then communicate that result off chip. Before this design was created, interfacing with ultra-low power chips was done either very slowly or with custom circuitry on both sides. The aim of this design was to use readily available hardware like an STM32 microcontroller and still be able to communicate with an ultra-low power device at reasonable speeds. Conventionally, I2C or SPI would be used for this task however the power budget restrictions make these protocols unfeasible. The pull-up resistors used in I2C consume power constantly with static leakage that can easily exceed the limited power supply of the low power systems an EH ALS may be included in. SPI requires at least 4 wires to interface between 2 devices, one of which is charged up by the secondary device.

Assuming this MISO wire has 20pF capacitance, if half of the entire chip power budget was used only for charging this wire, with the bits sent consisting of 0s half the time and 1s the other half then the maximum transfer speeds achievable with SPI would be 27.5kbps. Essentially, these existing protocols are not energy efficient enough to be used in realistic ultra-low power systems. MBus [4] was created to address these concerns but requires a considerable protocol overhead and a daisy chained topology that is incompatible with most existing devices.

Therefore, a novel communication bus and protocol has been developed based on I2C that consists of a push-pull clock and active open drain data wires shown in figure 2. This new active open drain configuration replaces pull-up resistors with an inverter connected to the main device. This inverter has much lower static leakage and is used to pull-up the data line. The secondary device can either leave this line high to send a 1 or pull it down to send a 0. The point is that the secondary device does not need to expend energy charging up the data line and so this new protocol is called Zero Power Communication (ZPC). It allows system designers to skew the energy requirement for communication away from the ultra-low power chip and towards a main device supplying both the clock and charging up the data line. An additional benefit to using an inverter as a pull-up mechanism is that it can vary the amount of current that flows through it on the fly. This gives the ability to change pull-up strength dynamically meaning

more control over the speed and power consumption of communication. Being able to dynamically change the communication speeds allows for more aggressive power domaining which is important in low-power applications. The ZPC block triggers entirely off the externally provided clock line which means that communication can happen at much faster speeds than the on-chip clock allows.
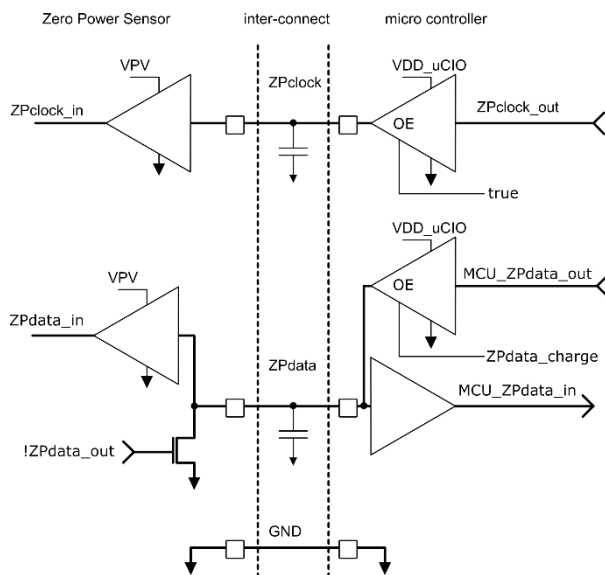

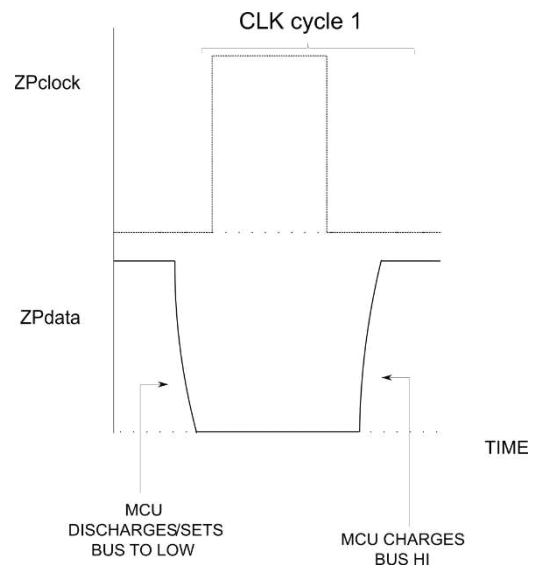Figure 2: Zero Power Communication Bus

Figure 3: Control bit timing

As shown in figure 3, if the data line is low on the positive edge of the clock, then the secondary device cannot communicate back and so this is defined as a control bit. The control bit is important to both structure the message sent but also define the lengths of fields within it. The novelty of the protocol is that these field lengths can change on a message-to-message basis based on how many successive control bits there are in different parts of the message: the first set of control bits define the length of the payload field, the next set define the length of the device address field, and the final set define the length of the register address field.

In the presented design, there are 2 modes of communication: addressing and non-addressing modes. In the addressing mode, all 3 of the payload, device address, and register address fields are present whereas in the non-addressing mode, the message only contains the payload field and the associated control bits. This is a way of reducing protocol overhead for the most common communication case for an EH ALS – reading out the ambient light measurement. Using the non-addressing mode only a single control bit needs to be sent to the sensor in order to read back an 8 bit ambient light measurement.

By using the control bits to indicate the lengths of each field provides an inherent error checking since both devices are informed of the field lengths before the message arrives. An error in communication cannot be detected using I2C without writing and reading back which consumes valuable power. The ability to change field lengths on a per message basis feeds into the aggressive power domaining now available to system designers: by changing what each number of successive control bits mean in terms of field lengths, common messages can be sent with optimal energy efficiency by resizing the message and having as few wasted bits as possible. Combined with control over how much current is used to pull-up the data line, this protocol promises the most flexibility in communication for the purpose of power efficiency.

RESULTS

At minimum irradiance ($0.33 W/m^2$) the analogue blocks have been simulated to cumulatively consume 17.9nW. The sampling frequency of the ADC at minimum irradiance is 18Hz with the maximum DNL and INL being 1.05 and 1.37 respectively. Ideally both figures would be below 1 but since there are no consecutive code losses, the ADC performance is acceptable. The digital errors are due to variation in the current reference generator affecting voltage reference generator and thus the operation of the

ADC	Architecture	Power Consumption (W)	Output Resolution (bits)	Sample Rate (Hz)	Energy per Code (pJ)
A 0.3V Biofuel-Cell-Powered Glucose/Lactate Biosensing System Employing a 180nW 64dB SNR Passive ΔΣ ADC [5]	Sigma-Delta	180n	11	3k	5.45
A Low-Power Incremental Delta-Sigma ADC for CMOS Image Sensors [6]	Sigma-Delta	29.5u	10	20M	0.15
A 100nW 10-bit 400S/s SAR ADC for Ultra Low-Power Bio-Sensing Applications [7]	SAR	100n	10	400	25.0
A 53-nW 9.1-ENOB 1-kS/s SAR ADC in 0.13- um CMOS for Medical Implant Devices [8]	SAR	53n	10	1k	5.30
SOLAS [9]	SAR	1.4n	6	20	11.67
MIMOSA [10]	Sigma-Delta	11.8n	8	36	40.97
AURORA (This work)	SAR	1.9n	8	20	11.88

Table 1 Comparison of EH and non-EH ADCs

ADC. This is due to the variable voltage supply from the PV cells and the aforementioned dependencies of the current and voltage references on this supply.

At minimum irradiance, the ZPC block consumes 84nW with a ZPClock frequency of 50kHz. This figure was found by using accurate power analysis tools on the place and routed digital netlist. In total, the presented design should consume around 102nW at minimum irradiance. The uncertainty of the final power consumption figure is due to the fact that high to low level shifters will need to be implemented between the blocks as the analogue circuitry is in the GO2 domain whereas the ZPC block is in GO1. The ZPC block has been implemented on an FPGA and has been able to communicate with an STM32 MCU with speeds up to 6.8kbps. The bit error rate of ZPC is <1E-6 achieved through validation testing. Finally, the power consumption of ZPC is 8.4pJ/bit which is half the energy consumption of MBus – the current state of the art communication protocol in terms of energy efficiency.

CONCLUSION

The presented work is an example of how an energy harvesting sensor can both conserve power through careful analogue design and interface with conventional devices that do not have such constrained power budgets. It is a promising start but the design needs to be manufactured and characterised before its usefulness can be determined. Mainly, the linearity of the ADC needs to be checked as well as the bit error rate of ZPC. Using ZPC, designers can now offload the energy requirement of communication away from ultra-low power devices towards devices with larger power budgets but the energy for communication still needs to come from somewhere. It remains to be seen if the overall system efficiency of ZPC is an improvement on current methods.

REFERENCES

[1]   F. Kaklin, J. M. Raynor and R. K. Henderson, "High Voltage Generation Using Deep Trench Isolated Photodiodes in a Back Side Illuminated Process," 2018 IEEE International Electron Devices Meeting (IEDM), San Francisco, CA, USA, 2018, pp. 32.2.1-32.2.4, doi: 10.1109/IEDM.2018.8614656.

[2]   P. Aphalo, "Photon Emission Spectra" 2009. [Online]. Available: https://www.mv.helsinki.fi/home/aphalo/photobio/data/spectra/

[3]   F. Kaklin, "Modelling and Simulation of an Energy Harvesting CMOS Image Sensor with Data Compression", PhD thesis, University of Edinburgh, Edinburgh, UK, 2022.

[4]   P. Pannuto et al., "MBus: An ultra-low power interconnect bus for next generation nanopower systems," 2015 ACM/IEEE 42nd Annual International Symposium on Computer Architecture (ISCA), Portland, OR, USA, 2015, pp. 629-641, doi: 10.1145/2749469.2750376.

[5]   A. Fazli Yeknami, X. Wang, S. Imani, A. Nikoofard, I. Jeerapan, J. Wang, P. Mercier, "A 0.3V biofuel-cell-powered glucose/lactate biosensing system employing a 180nW 64dB SNR passive $\delta\varsigma$ ADC and a 920MHz wireless transmitter," in IEEE International Solid - State Circuits Conference - (ISSCC), San Francisco, USA, 2018.

[6]   I. Lee, B. Kim, B. Lee, "A Low-Power Incremental Delta–Sigma ADC for CMOS Image Sensors," IEEE Transactions on Circuits and Systems II: Express Briefs , vol. 63, no. 4, pp. 371-375, 2015.

[7]   H. Franca, M. Ataei, A. Boegli, P. Farine, "A 100nW 10-bit 400S/s SAR ADC for ultra low-power bio-sensing applications," in 6th International Conference on Informatics, Electronics and Vision & 2017 7th International Symposium in Computational Medical and Health Technology (ICIEV-ISCMHT), Himeji, Japan, 2017.

[8]   D. Zhang, A. Bhide, A. Alvandpour, "A 53-nW 9.1-ENOB 1-kS/s SAR ADC in 0.13-μm CMOS for Medical Implant Devices," IEEE Journal of Solid-State Circuits , vol. 47, no. 7, pp. 1585-1593, 2012.

[9]   S. Ball, "Design of an Energy Harvesting Ambient Light Sensor with 6-Bit Digital Output," MEng Individual Report, University of Glasgow, Glasgow, UK, 2021.

[10]   I. Todorova, "Design of a 3D Stacked Energy Harvesting Ambient Light Sensor with 8-Bit Digital Output," MEng Individual Project, University of Glasgow, Glasgow, UK, 2022.

An Efficient Direct Time-of-Flight (dToF) LiDAR System Based on High Resolution SPAD Array

Tze Ching Fung*, Chunji Wang, Hongyu Wang, Alfonso Cesar Barredo Albason,
Radwanul Hasan Siddique and Yibing M. Wang

Meta Vision Lab, Samsung Semiconductor Inc., Pasadena CA, USA
*Email: richard.fung@samsung.com

Abstract—**We proposed a LiDAR system for mobile application that is power efficient and tolerates high ambient. The system has 940nm scanning laser sources and a 192×144 single-photon avalanche diode (SPAD) array with on-chip time correlated histogramming. The adaptive single-pass histogramming architecture saves power by staying at coarse resolution mode when returned laser pulses are weak. New signal processing method allows a depth resolution beyond coarse bin size. The system is able to measure depth images at 30 fps, with up to 10 meters range, 1% range error, and only consumes 12 mW optical power.**

Keywords—**LiDAR, SPAD, dTOF, adaptive, single-pass, histogram**

I. INTRODUCTION

Depth sensing has become a salient feature in applications that involve environmental interactions, such as augmented reality (AR), self-driving vehicles, and security surveillance. Among common ranging technologies like radar and ultrasonic, Light Detection and Ranging (LiDAR) sensors have advantages of long range, high range resolution, and high spatial resolution. LiDAR resolves depth via measuring the time of flight (ToF) of light between the sensor and the detected object. According to different measurement principles, it is often categorized either as direct time of flight (dToF) [1] or indirect time of flight (iToF) [2] types. dToF is of particular interest due to its immunity to multi-path echo and the ease of augmenting it with existing vision systems to form a 3D image solution [3].

A dToF LiDAR sensor needs to reject the background light (aka ambient) from the transmitted light by repetitive measurement. To detect returned laser photons from all photons, a histogram is typically used to bin the photon timestamps, and the peak of the histogram usually indicates returned laser photons. It can be quite challenging to design a system that strikes the right balance between power efficiency, ambient tolerance, range, and range resolution. Increasing light source intensity or reducing histogram bin size often warrant better performance; but the increase in power consumption and memory may negate the benefit. A two-pass solution has been proposed [4] to mitigate this issue. It first uses a coarse histogram to estimate depth in coarse resolution, then build a new fine histogram around the estimated coarse depth, so that a fine resolution depth is obtained. Nonetheless, the signals in the first pass are discarded before the second pass, making no contribution to improve range resolution. The impact of signal loss (or waste) is especially pronounce at long range measurement where returned photons are usually fewer and therefore more valuable.

In this paper, we present a design of a LiDAR system using 192 × 144 SPAD array. The system features an adaptive single pass histogramming architecture that allows the histogram to adaptively switch from coarse mode to fine mode depending on returned laser signals strength. Together with novel signal processing, the system is capable to detect low albedo (10%) objects with high resolution at 10 meters under full sun outdoor, at 30 frames per second (fps), at low power consumption.

II. SYSTEM ARCHITECTURE

The system architecture is illustrated in Figure 1. We adopted scanning LiDAR scheme to take advantage of its higher signal-to-noise ratio (SNR) than non-scanning ones (e.g. Flash LiDAR) [5]. Different types of laser beam scanners have been investigated for LiDAR application; among those, solid-state scanners are often preferred due to no moving parts, fast response time, potential for a compact design, and low cost [6]. Especially recent advancement in combining laser diode array with diffractive optics element (DOE)
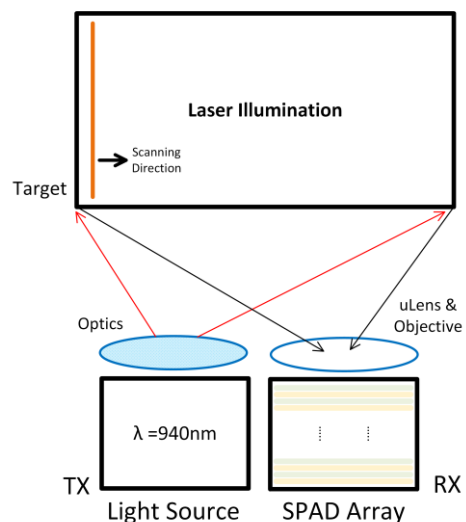

Figure 1. Proposed LiDAR system in this work including a 940nm infrared light source as transmitter (TX) and an addressable SPAD array as receiver (RX).
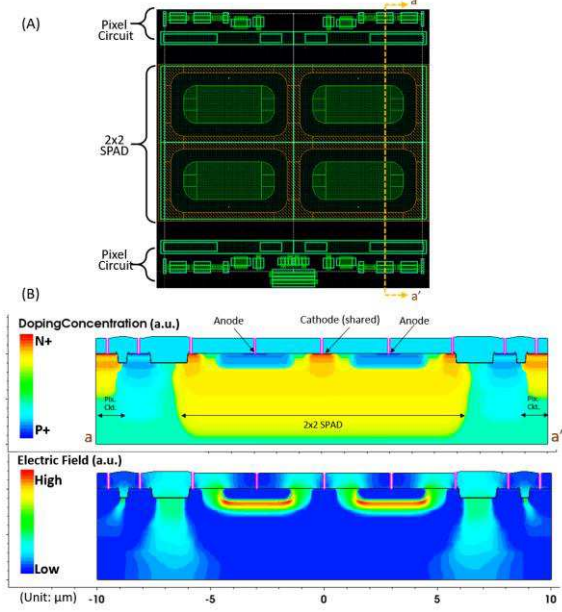
Figure 2. (A) The 20μm pixel layout, and (B) cross-sectional view in doping concentration and electrical filed distribution by TCAD simulation.
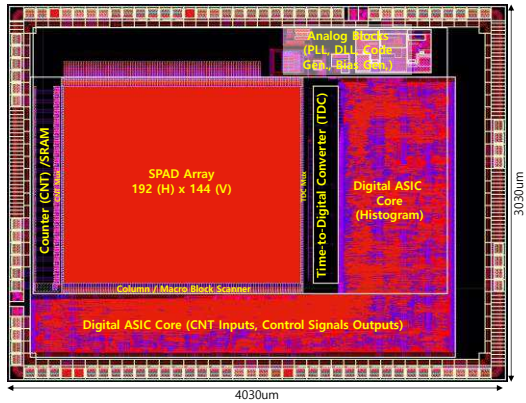

Figure 3. Layout of 90nm BSI SPAD sensor chip (RX).

makes them very attractive in LiDAR systems for consumer products [7].

The TX consists of a laser scanner with optical elements illuminating the field with line patterns at an infrared wavelength of 940 nm. Due to the optical disparity between the laser patterns and the light signals reflected back from the target, the interrogation pattern spacing cannot be uniform throughout. We designed the SPAD array in the receiver (RX) with a flexible addressing scheme to avoid missing reflected light signals, which is achieved by dividing the SPAD array into different regions, such that the scanning and addressing patterns within each regions can be set independently. As a result, the RX scanning patterns can be tuned to match perfectly with the changing reflected light locations to guarantee the capture of returned signal photons.

As will be described in detail below, the RX chip has also counter (CNT) and time-to-digital converter (TDC) circuits placed on both sides of the array
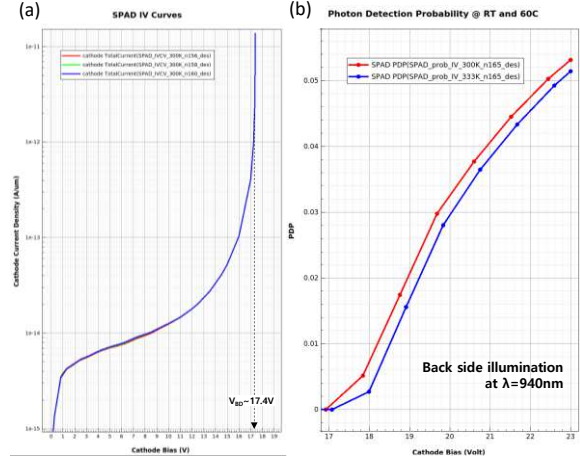
Figure 4. TCAD simulation of (a) SPAD reverse bias I-V curve and (b) PDP as a function of cathode bias.

allowing SPAD voltage pulses to be processed immediately after photons being detected. In addition, with on-chip ASIC supporting full histogram building and readout, the proposed LiDAR system is capable of real-time depth sensing at 30 fps.

III.   SPAD SENSOR CHIP AND OPERATION

We designed the RX with all the essential blocks to conduct the dTOF operation fully on-chip. Figure 5 illustrates the SPAD sensor chip architecture, signal paths, and the 2×2 SPAD pixel circuit. A SPAD typically operates in avalanche breakdown region with a reverse bias slightly exceeding its breakdown voltage (i.e. $V_{bias} = V_{bd} + V_{ex}$) with V_{ex} being the excess voltage. This means SPAD has very high signal gain and can detect single-photon events. To reset an excited SPAD back to its stand-by state, a quenching transistor is connected at the anode (aka passive quenching) [8]. This makes a compact SPAD design difficult to achieve. Recently virtual guard ring based SPAD designs have shown to be promising in achieving small pitch [9]. We built upon a similar concept, and further improved the fill factor (to 19.6%) by sharing one common cathode and output circuit with four (2×2) neighboring SPADs in a 20 μm pixel (refer to Figure 2). The SPAD has a pitch of 10 μm, and we designed the breakdown voltage (V_{bd}) at 17.4V. The four SPAD output channels are coupled through an OR gate as a unified pixel output to save circuit area and improve fill factor. Signal processing of the combined pulses will be discussed in detail in Section IV. The SPAD sensor chip is fabricated by Samsung's custom 90 nm back-side illumination (BSI) image sensor process. TCAD simulations confirmed the present design can achieve photo detection probability (PDP) of 3% at 940nm with an excess voltage (Vex) of 2.6V (Fig.4). By optimizing device structure, incorporating metalens to improve effective fill factor, and nanostructured thin-film to enhance light absorption, we estimate a greater than 3-folds improvement with epi-thickness of 3um, or photo detection efficiency (PDE) above 10% in final designs.

As shown in Figure 6, the dTOF operational sequence can be separated into two parts: locating
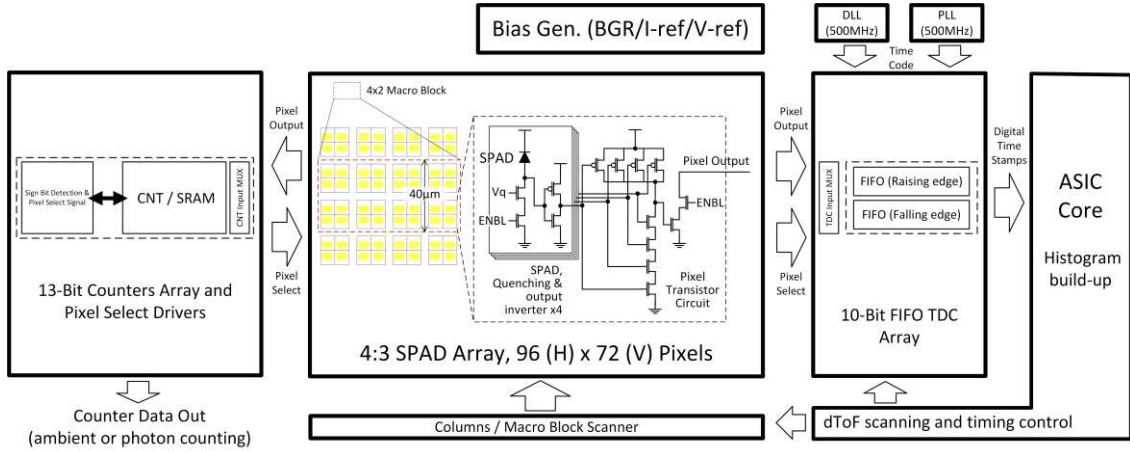
Figure 5. Sensor chip (RX) architecture. The SPAD array has a total of 96×72 (2×2 SPAD) pixels and the inset illustrates the 2×2 SPAD pixel circuit. The 13-bit counters (CNT) and dual 10-bit time-to-digital converters (TDC) are located on the left and right side of array, respectively.
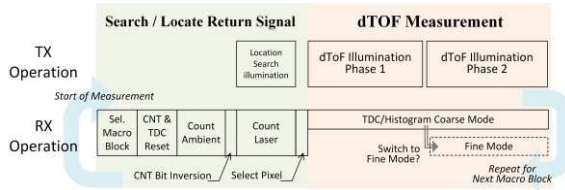

Figure 6. dToF operational sequence of one line scan. The sequence repeated itself until the TX scans through the entire scene.

reflective laser signals followed by repetitive dTOF measurement cycles.

At the beginning of return signal search, new groups of pixels are enabled by signals from column/ macro block scanner (ref. Figure 5, located at the bottom of array). As can be seen in Figure 5, eight (4×2) neighboring pixels are addressed together as a macro block, so the returned laser photons can be captured regardless of disparity. Once the macro block is selected, the 13-bits CNT array first counts the ambient photons for all pixels within the block. CNT bit inversion (2's compliment) is then performed followed by another round of photon counting under laser illumination (labeled "location search" in Figure 6). Exact return laser signal location can then be determined by detecting which CNT (hence pixel) first has its sign bit flipped to positive. Such operation allows us to design the array with just a simple upward CNT instead of a more complicated bidirectional (up/down) one. At the end of this phase, CNT array sends out the pixel select signals which drive the TDC input MUX. Therefore, only the pixel that received the highest amount of photons is selected for dToF measurement.

During the dTOF measurement phase, digital pulse trains from SPAD array are processed by high precision TDC capable of resolving time stamps down to 250 ps. The total number N of laser pulses to be projected is determined by the maximum range to be measured. Midway through the measurement, the ASIC logic analyzes histogram from the first M cycles and determines if it should switch to the fine bin mode. This dTOF operation is repeated until it scans through the entire SPAD array.

IV. SIGNAL PROCESSING

Since the laser pulse is quite wide (e.g. w = 4 ns, about 0.6 meter in distance), edge triggering is required to achieve the desired resolution (1% of range). We designed the signal processing pipeline (i.e. TDC) to accurately measure the time stamps of rising and falling edges of returned pulses. The signal processing pipeline consists of pulse collection recovery, histogramming, and post-processing.

When two or more SPADs give pulses simultaneously, they merge into a single pulse on the readout channel (we call this process "pulse collision" or "event collision"), leading to signal loss. Pulse collision happens more frequently if a single SPAD pulse is longer. When pulses collide, only the first pulse leading edge and the last pulse trailing edge survive; other edges are lost.

This is similar to photon pile-up [10] in that later photons get masked by earlier photons. Pulse collision compounds with pile-up, distorts the photon counts and introduces depth measurement error. To alleviate this signal loss, we developed pulse collision recovery algorithm (Figure 7A) to recover most of the lost signal. To do this, TDC records both the leading and trailing edges timestamps (t_1 and t_2) of a pulse. If $t_2 - t_1$ is greater than the single pulse width (w_S) in a single SPAD channel (obtained via calibration), then we can recover one more event timestamp at $t_2 - w_S$. We show an example in Figure 7A, in which three SPADs detect in total 6 photons, but the merged channel only shows 3 pulses. The collision recovery algorithm recovers 2 out of 3 lost photon-detection events. Even though not all photons are recovered, simulation shows in Figure 7B that this recovery algorithm leads to result approaching to the case without collision at all. This is because the probability of n pulses colliding decreases dramatically with n, so for the majority of collision cases, those of two pulses colliding, are fully recovered.
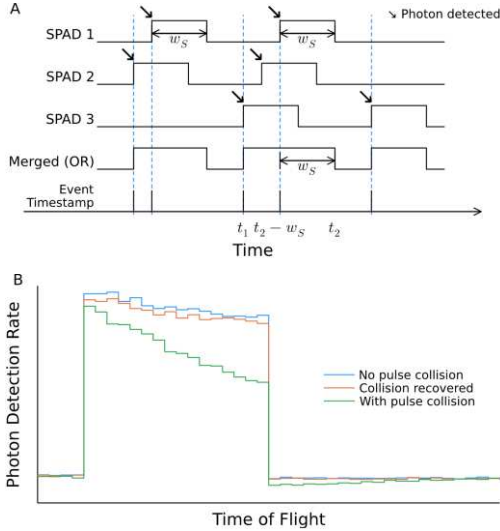
Figure 7. (A) Collision recovery algorithm example showing three SPAD channels merging into one channel. (B) Simulated histogram with pulse collision and its recovery algorithm. Target albedo is 0.5 at 7 meters, under half of outdoor full-sun ambient.

To obtain precise TOF from the fine histogram, it is necessary to apply a digital filter to enhance the peak. To avoid pile-up induced issues, we developed a conditional FIR filter for accurate peak enhancement. Figure 8D illustrates an example by simulation. The resolved TOF matches well to the peak location of the filtered histogram.

V. PERFORMANCE ESTIMATION AND CONCLUSION

We numerically simulated the system with Monte Carlo method to estimate system measurement error. As shown in Figure 9, the error keeps within 1% of range, except at extremely close range below 1.5 m, where the error cannot be reduced beyond fine bin resolution (250 ps, equivalently 3.75 cm). Table 1 gives a summary of the system specs. In conclusion, we reported a complete LiDAR system design including TX, RX, on-chip histogramming and data processing. The TX is based on laser scanner illuminating at 940nm. The RX has a SPAD array with a flexible addressing scheme for disparity-agnostic return signal
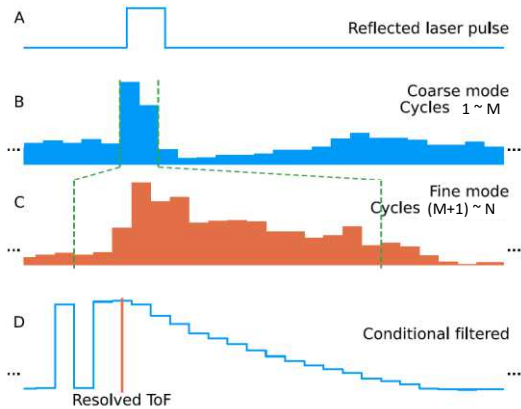
Figure 8. Adaptive single pass histogram, switching from coarse to fine mode in the middle of the N cycles. (D) shows a conditional FIR filter.
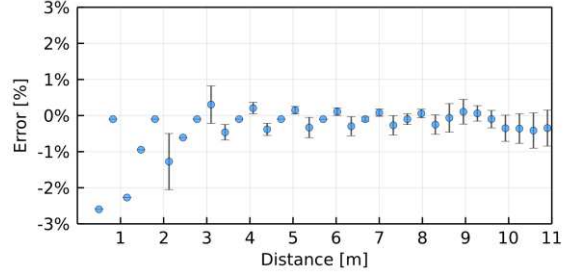
Figure 9. Simulated measurement error. Each data point and represents statistic of 100 repeated measures. Target albedo is randomly sampled from 0.1 to 0.8; ambient level is randomly sampled from 0%~100% full sun ambient.

locating. On-chip adaptive single-pass histogram shows superior performance compared to time-gated solutions [4] under limited optical power budget. The system is capable of capturing depth images at 30 fps at 12 mW optical power. The system has a range of up to 10 meters, and depth error is kept within 1%. The system is well-balanced between energy efficiency, range, range resolution and ambient tolerance, making it suitable for mobile ranging application.

TABLE I.   SYSTEM CHARACTERISTIC (SIMULATION)

System Specification	Value
Avergae Opt. Power	12mW
Maximum Range	10 meters
Range Precision	< 1% × range
Frame Rate	30 fps
TDC Resolution	250 psec

REFERENCES

[1] Gongbo Chen, Christian Wiede, and Rainer Kokozinski, "Data Processing Approaches on SPAD-Based d-TOF LiDAR Systems: A Review," IEEE Sensors Journal, 21(5):5656–5667, Mar.

[2] Cyrus Bamji, John Godbaz, Minseok Oh, Swati Mehta, Andrew Payne, Sergio Ortiz, Satyadev Nagaraja, Travis Perry, and Barry Thompson, "A review of indirect time-of-flight technologies," IEEE Trans. Electron Devices, 2022.–73.

[3] I. Gyongy, N. A. W. Dutton and R. K. Henderson, "Direct Time-of-Flight Single-Photon Imaging," in IEEE TED, vol. 69, no. 6, pp. 2794-2805, June 2022

[4] AK Sharma, A Laflaquiere, GA Agranov, G Rosenblum, and S Mandai, "Spad array with gated histogram construction," US Patent 20170052065 A, 1, 2017.

[5] Dingkang Wang, Connor Watkins, and Huikai Xie, "Mems mirrors for lidar: A review," Micromachines, 11(5), 2020.

[6] Thinal Raj, Fazida Hanim Hashim, Aqilah Baseri Huddin, Mohd Faisal Ibrahim, and Aini Hussain, "A survey on lidar scanning mechanisms," Electronics, 9(5), 2020.

[7] Yulong An, Yanmei Zhang, Haichao Guo, and Jing Wan, " Compressive sensing-based threedimensional laser imaging with dual illumination," IEEE Access, 7:25708–25717, 2019.

[8] Andrea Gallivanoni, Ivan Rech, and Massimo Ghioni," Progress in quenching circuits for single photon avalanche diodes," IEEE Trans. on Nuclear Science, 57(6):3815–3826, 2010.

[9] Tomer Leitner et al., "Measurements and simulations of low dark count rate single photon avalanche diode device in a low voltage 180-nm cmos image sensor technology, "IEEE Trans. on Electron Devices, 60(6):1982–1988, 2013.

[10] P B Coates, "The correction for photon 'pile-up' in the measurement of radiative lifetimes," Journal of Physics E: Scientific Instruments, 1(8):878–879, Aug. 1968

SLIM: Small and Learnable Image Signal Processing Module for CMOS and Quanta Image Sensors

Stanley H. Chan[1], Yiheng Chi[1], and Preston Rahim[1]

[1] DeepLux Technology Inc., West Lafayette, IN 47906, USA
{stanley.chan, yiheng.chi, preston.rahim}@deeplux.tech

Abstract. While multibit Quanta Image Sensors (QIS) today have demonstrated a superior sub-electron read noise characteristic, at extreme photon-limited conditions they still face the fundamental photon shot noise problem. The image signal processing (ISP) unit of today's multibit QIS is largely identical to those used for CMOS image sensors. In extreme photon-limited conditions, these physics-based ISP struggle to generate high-quality images. Deep learning methods are seen as the potential solution to overcome the low-light bottleneck, but existing neural networks are too large to fit into any camera products. In this paper, we present a learning-based ISP where key components are replaced by a lightweight neural network followed by traditional physics-based filtering steps. The proposed ISP, known as the Small and Learnable ISP Module (SLIM), allows us to jointly demosaick and denoise images at a photon level as low as 1 photon per pixel where traditional ISP fails.

1. Introduction

Single-bit and multi-bit Quanta Image Sensors (QIS) have delivered promising low-light image capturing results with sub-electron read noise. However, as the total amount of photon flux drops, QIS will eventually encounter the fundamental photon shot noise limit. The mainstream image signal processing (ISP) units today have limited capability of handling an excessive amount of shot noise. Many of them are still using the classical signal processing techniques based on engineered heuristics. Recent advancements in deep learning has generated a significant interest in the ISP community where people start to consider upgrading these traditional ISP to a learning-based ISP. Yet, the complexity of learning-based ISP (especially those using deep neural networks) is so high that even a high-end mobile phone processor only performs such operations occasionally when a user needs to restore a single photograph. For mid-grade and low-end products such as laptops, medical devices, cars, and household appliances, pushing artificial intelligence to ISP becomes a pressing demand that will continue for the coming decade.

In this paper, we present an algorithmic solution for two critical steps in the ISP pipeline: low-light denoising and low-light demosaicking. Compared to the traditional ISP and deep learning based ISP, our solution can be seen as a middle-ground solution that balances performance and complexity:

- Compared to traditional ISP that performs rule-based demosaicking and linear filtering (usually edge-aware weighted averaging and median filtering), our proposed solution uses a few shallow layers of neurons to extract high and low level features across different scales. The denoising is performed by a chain of new procedures to construct and select denoising filters. These new procedures alleviate the limitations of traditional ISPs which often fail to identify edges and texture when the input is corrupted by heavy noise.

- Compared to deep-learning based ISP such as [1] that requires training large models end-to-end, our proposed solution is significantly more light-weight. We use simple convolutions and shallow layers of neurons to perform most of the tasks, in contrast to complex models such as vision transformers and self-attention. Moreover, since our design is based on the traditional pipeline where parts are modularized, it makes debugging and interpretation easier.

2. Small and Learnable ISP Module (SLIM)

Figure 1 illustrates the schematic diagrams of a typical ISP and our proposed ISP. The input to the ISP is the Bayer color filter array pattern assuming that the standard pre-processing steps are completed (e.g., gray-level offset, pixel response non-uniformity calibration, dead pixel removal, etc.) The focus of our work is the demosaicking and denoising steps in the raw domain. The output is sent to a downstream ISP module for additional processing of color and edge. In the proposed pipeline, we replace several key steps of the denoising process by learning based methods:
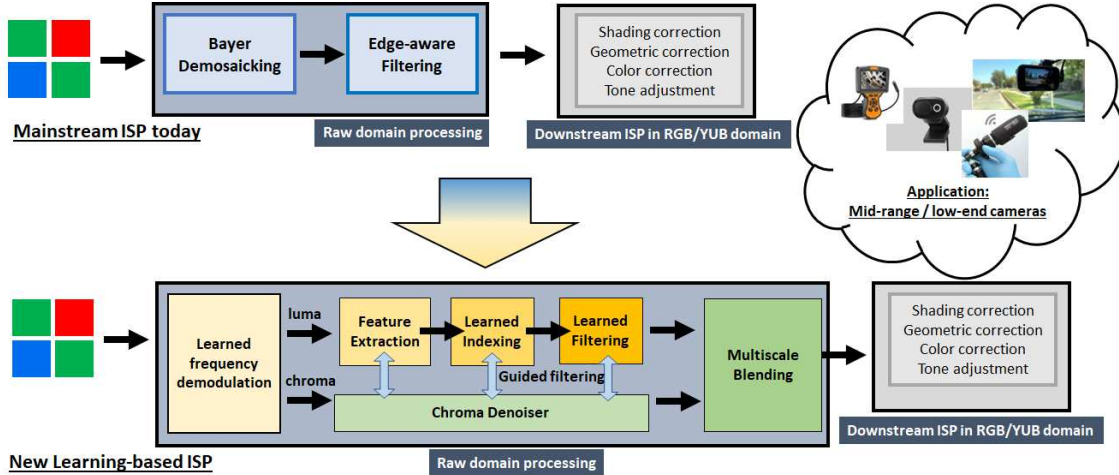
Figure 1: Schematic diagram of the proposed ISP compared to a typical ISP. In our proposed ISP, we replace several key steps of the traditional ISP with learning-based modules. These modules are light-weight shallow neural networks that are trained independently of each other. For end applications, we target mid-grade to low-end cameras such as laptops, dash-cams, endoscopes, and inspection cameras.

- **Learnable frequency selection**. We use a frequency selection module [2] to solve the Bayer color filter array (CFA) demosaicing problem in SLIM. When light goes through the CFA, the color channels are modulated by carrier signals with known frequencies. Given this mosaiced signal, the full-resolution color information can be recovered by signal demodulation operations. Traditional methods use linear demodulation schemes which are not robust to severe noise under low lighting conditions [4]. Recent deep-learning based demosaicing methods usually require excessive computing resources, memory, and execution time, while they often ignore the physics of CFAs. Instead, we propose a learnable light-weight frequency selection module. This module is developed upon the physics of the Bayer pattern CFA and is adaptive to various signal-to-noise ratios under different lighting conditions.

- **Feature extraction**. SLIM extract features from the luma channel of the demodulated image signal before denoising it. Such nonlinear feature extraction is crucial to the later stage of denoising because it accumulates spatial information, which helps reconstruct the image when the input is contaminated by heavy shot noise at low light.

- **Learned indexing**. Instead of using a single filter or a selected filter from a collection, our proposed SLIM uses a learned indexing scheme to calculate a combination of multiple filters and their corresponding strengths (weights) from the extracted features. Such indexing is continuous and back-propagatable. With this learned indexing, SLIM composes sophisticated and nonlinear filters similar to deep networks do, but the computational complexity of SLIM is much lower so it can be empowered by edge devices.

- **Learned filtering**. Like many data-driven methods today, SLIM learns the image filters from vast image data. The training data are synthesized using realistic image formation models at various light levels. When deployed, processing image signals with these predetermined learned filters can save much memory compared to generating full-resolution images using deep restoration networks in one pass.

2

- **Chroma-Luma decoupling**. The SNR of luma signal is higher than chroma signals due to their derivations. Therefore, we distribute most computations towards luma channel processing. Afterwards, we use the luma features and indices to guide the denoising of the chroma channels.

- **Multi-scale blending**. A pyramid-shape multi-scale structure has been demonstrated effective in image processing literature. SLIM performs a two-level multi-scale blending to perform global-then-local image denoising.

Method	# Parameters	Description
Deep-Learning ISP		
MIRNet 2020 [9]	32,787,000	Denoising or super-resolution or image enhancement
Restormer 2022 [8]	26,127,000	Denoising or deraining or deblurring
DRUNet 2021 [10]	32,640,000	Denoising or demosaicing
PyNet 2020 [5]	47,554,000	ISP (demosaicking + denoising + white balance + color correction)
Traditional Filter-Bank ISP		
RAISR 2016 [7]	26,000	Superresolution
BLADE 2017$^+$ [3]	28,000	Denoising
Proposed Learned Physics-based ISP		
SLIM (proposed)	126,000	ISP (demosaicking + denoising)

Table 1: Comparison between the number of parameters used in various ISPs. The number of parameters is measured in terms of the size of the filters and the number of filters. $^+$ Existing mainstream ISP are rule-based edge-aware filters. While they do not explicitly build the filter bank, the number of filters they use are in the same order of magnitude compared to BLADE.

Results and Conclusion

Table 1 shows the number of parameters used by existing deep learning models and SLIM. Figure 2 shows a comparison between a proprietary ISP on one of the mid-grade cameras and the proposed ISP. The input Bayer images are simulated at photon levels from 1 ppp to 100 ppp, assuming a 0.19e- read noise, 0.02e-/s dark current, 12-bit analog-digital converter, 80% quantum efficiency, and a uniform sensor response. We train the learning-based modules *independently* using simulated data across a wide range of photon levels. Such an independence implies that the whole ISP does *not* need to be trained end-to-end. Thus, updating one module does not interfere with another module, hence making the debugging at the same level as a traditional ISP (and much more convenient than a deep-learning ISP). To summarize, SLIM demonstrates the potential as a viable solution for the next generation learning-based ISP where the photon level is low.

References

[1] Y. Chi, A. Gnanasambandam, V. Koltun, and S. H. Chan. Dynamic low-light imaging with quanta image sensors. In *Proc. European Conf. Computer Vision*, pages 122–138, 2020.

[2] O. A. Elgendy, A. Gnanasambandam, S. H. Chan, and J. Ma. Low-light demosaicking and denoising for small pixels using learned frequency selection. *IEEE Trans. Computational Imaging*, 7:137–150, 2021.

[3] P. Getreuer, I. Garcia-Dorado, J. Isidoro, S. Choi, F. Ong, and P. Milanfar. BLADE: Filter learning for general purpose computational photography. In *IEEE International Conference on Computational Photography (ICCP)*, pages 1–11, 2018.

[4] A. Gnanasambandam, O. Elgendy, J. Ma, and S. H. Chan. Megapixel photon-counting color imaging using quanta image sensor. *OSA Optics Express*, 27(12):17298–17310, 2019.

[5] B.-H. Kim, J. Song, J. C. Ye, and J. Baek. Pynet-ca: enhanced pynet with channel attention for end-to-end mobile image signal processing. In *Proc. European Conf. Computer Vision*, pages 202–212, 2020.

[6] J. Ma, D. Zhang, O. A. Elgendy, and S. Masoodian. A 0.19e- rms read noise 16.7mpixel stacked quanta image sensor with 1.1 um-pitch backside illuminated pixels. *IEEE Electron Device Letters*, 42(6):891–894, 2021.

[7] Y. Romano, J. Isidoro, and P. Milanfar. RAISR: Rapid and accurate image super resolution. *IEEE Trans. Computational Imaging*, 3(1):110–125, 2016.

[8] S. W. Zamir, A. Arora, S. Khan, M. Hayat, F. S. Khan, and M.-H. Yang. Restormer: Efficient transformer for high-resolution image restoration. In *Proc. IEEE/CVF Conf. Computer Vision and Pattern Recognition*, pages 5728–5739, 2022.

[9] S. W. Zamir, A. Arora, S. Khan, M. Hayat, F. S. Khan, M.-H. Yang, and L. Shao. Learning enriched features for real image restoration and enhancement. In *Proc. European Conf Computer Vision*, pages 492–511, 2020.

[10] K. Zhang, Y. Li, W. Zuo, L. Zhang, L. Van Gool, and R. Timofte. Plug-and-play image restoration with deep denoiser prior. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 44(10):6360–6376, 2021.
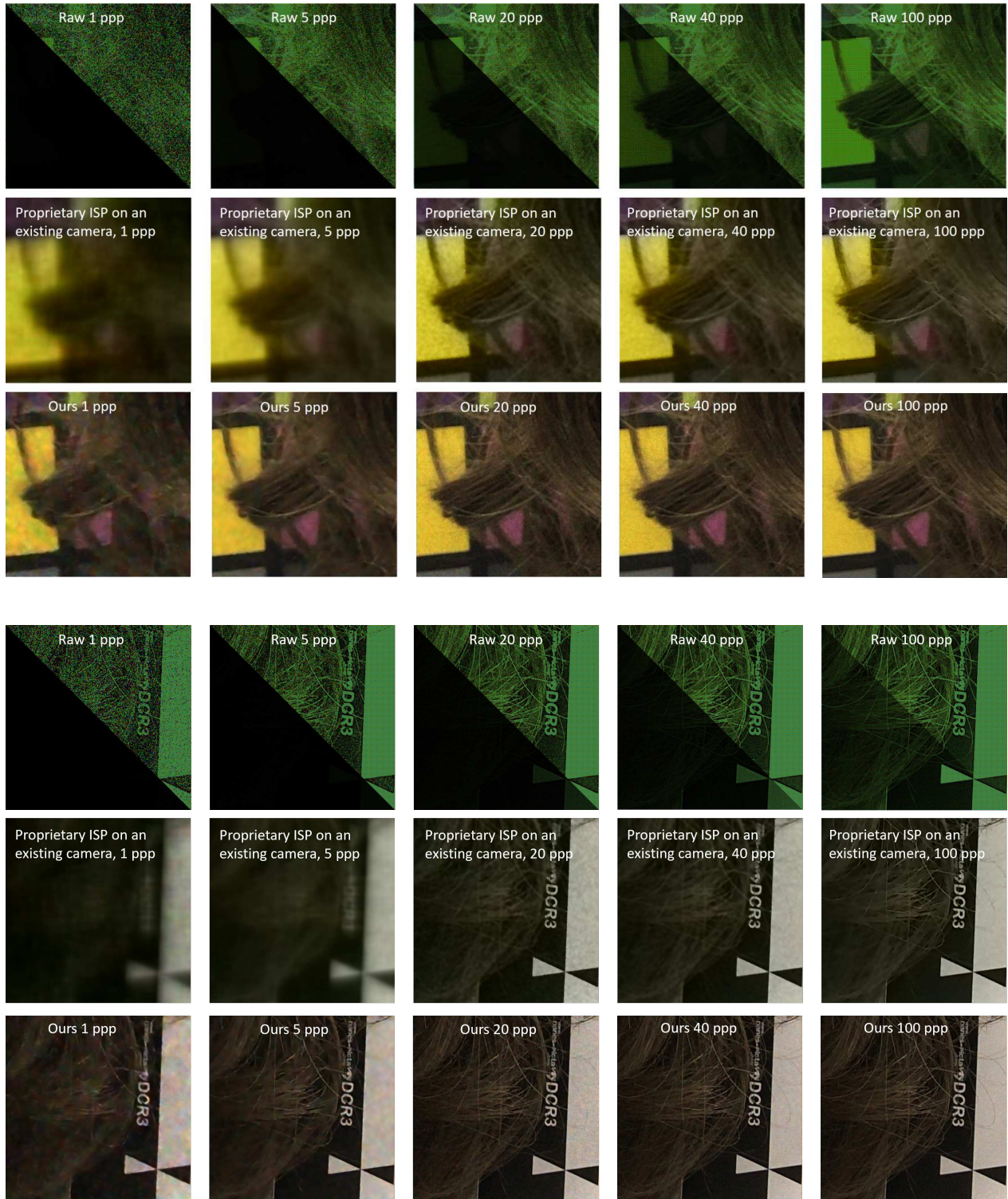
Figure 2: Denoising and demosaicking results of our proposed ISP, compared to a proprietary ISP used in existing image sensors. The noise level is indicated by the number of photons per pixel (ppp). The noise model follows a published QIS specification [6].

4

Cyber Security for CMOS Image Sensors

Boyd Fowler, Wenshou Chen and Kevin Johnson

OmniVision Technologies, Email: boyd.fowler@ovt.com / Tel.:+1-408-653-2309

Abstract—**This paper describes the current state of the art in cyber security for CMOS images sensors. It also shows some of the limitations of this technology, such as unidirectional certificate exchange and incomplete message authentication (MAC) of the image. Then an architecture is proposed that can overcome some of these limitations and improve data and command security. We also describe a framework for analyzing the security of these systems and use it to bound the security of both architectures described in this paper.**

I. INTRODUCTION

Cyber security is critical for protecting image data but it is a relatively new feature in image sensors. Modern cyber-security systems use standard algorithms in provably secure frameworks [1]. An example of this is transport layer security (TLS) used by HTTPS for internet data transmission. In addition to a provably secure framework, an efficient hardware implementation is also required to accelerate these algorithms, reduce power dissipation and to resist side channel attacks [2].

II. CURRENT CYBER SECURITY SYSTEMS

Figure 1 shows a typical cyber security implementation with an image sensor connected to an ASIC. In this implementation two methods are used to secure the system. The first technique uses a cryptographically signed certificate to authenticate the identity of the image sensor. The second is a message authentication code (MAC) used to guarantee the integrity of the image sensor data. The boot-up process between the ASIC and the image sensor starts by having the image sensor transmit a signed certificate, such as an X509.v3 [3], to the ASIC. Then the ASIC uses a global public key, stored in the ASICs non-volatile memory, to verify the certificate's signature is from a trusted source, such as the image sensor vendor. If the signature is valid then it loads the image sensor public key from the certificate into the asymmetric encryption engine. Then the ASIC creates a random secret key, using a pseudo random number generator (PRG), for the MAC process. This key is then asymmetrically encrypted using the sensor's public key and it is transmitted to the image sensor. The image sensor then decrypts the secret MAC key and loads it into the MAC hardware. Then a NONCE, a number that is only used once, is randomly generated in the image sensor and supplied to the MAC hardware and the image data is processed. A unique NONCE is needed for each image to make the process secure. Finally the image, the NONCE and the unique MAC tag are output and sent to the ASIC. After N images are transmitted from the sensor to the ASIC, a new secret key for the MAC process must be generated by the ASIC and sent to the image sensor to keep the system secure. N is typically a function of the systems susceptibility to side channel based key recovery attacks.

The system described in Figure 1 has many security limitations including an ASIC that can be compromised or bypassed, unencrypted image sensor data, and insecure image sensor control. In addition, due to limited computational resources often only a portion of the image is actually used in the MACing process. Although the part of image that is MACed is often randomized using the secret key this still leads to very poor security.

In order to understand the security of this system we need a few definitions. First we define a chosen plaintext attack (CPA) game for a MAC algorithm $\mathcal{I} = (S, V)$, where S is the MAC signing algorithm and V is the verification algorithm. This game is shown in Figure 2. It starts by having the challenger (image sensor) create a secret key k from key space $\mathcal{K}$ ($k \xleftarrow{R} \mathcal{K}$). Then the adversary $\mathcal{A}$ (the attacker) sends a set of messages $(m_0, m_1, m_2 \dots m_{q-1})$ from the message space $\mathcal{M}$ to the challenger. The challenges creates a tag $t_i \leftarrow S(k, m_i)$ from the tag space $\mathcal{T}$ for each received message and sends it back to the adversary. Finally the adversary tries to create a valid message tag pair (m_q, t_q), where $m_q \notin (m_0, m_1, m_2, \dots m_{q-1})$. If the adversary succeeds then the message tag pair is an existential forgery. The security, or advantage, of the MAC is the probability that the adversary creates an existential forgery ($MACadv[\mathcal{A}, \mathcal{I}]$). For practical cyber systems we need $MACadv[\mathcal{A}, \mathcal{I}]$ to be negligible. The definition of negligible does depend on the attacker but it is usually $< 2^{-90}$.

Using the CPA game above, assume that only a randomly selected fraction x/y of the image is MACed, where x is the number of MACed pixels and y is the total number of pixels. We will call this MAC algorithm $\mathcal{I}'$. In this scenario the attacker can send a single message to the challenger and receive a corresponding tag. Then the attacker can change a single pixel in the image and return the modified image and the received tag. The probability of winning this game is

$$MACadv[\mathcal{A}, \mathcal{I}'] \leq MACadv[\mathcal{B}_\mathcal{I}, \mathcal{I}] + (1 - x/y) \quad (1)$$

where $MACadv[\mathcal{B}_\mathcal{I}, \mathcal{I}]$ is the probability of winning the game assuming every pixel is MACed and $(1 - x/y)$ is the probability that the pixel selected by the attacker is not in the set of pixels that were MACed. Note that $\mathcal{B}_\mathcal{I}$ is a sub adversary of $\mathcal{A}$. Even if $x = y - 1$ and $y \sim 2^{27}$ the MAC advantage is $\gg 2^{-90}$.

Now we define a chosen cipher text attack (CCA) game. Given a cipher $\mathcal{E} = (E, D)$, where E is encryption algorithm and D is the decryption algorithm, defined over a key space $\mathcal{K}$, a message space $\mathcal{M}$ and a cipher-text space $\mathcal{C}$, the game

starts with the challenger randomly selecting a key $k \xleftarrow{R} \mathcal{K}$ and a binary value $b \in \{0,1\}$. Then the adversary $\mathcal{A}$ makes a series of queries to the challenger. Each query is either an encryption query or a decryption query. An encryption query consists of having the adversary send two messages of the same length $(m_{i0}, m_{i1}) \in \mathcal{M}^2$ to the challenger, then the challenger encrypts message $c_i \leftarrow E(k, m_{ib})$ and returns the cipher-text c_i to the adversary. A decryption query consists of having the adversary send cipher-text $c_j \in \mathcal{C}$ that is not the response of any of the encryption queries. The challenger then computes $m_j \leftarrow D(k, c_j)$ and returns the decrypted message to the adversary. The adversary can initiate as many of these queries as necessary in any order. Then at the end of the game the adversary computes the value of b. $\mathcal{A}$'s advantage with respect to $\mathcal{E}$ is

$$CCAadv[\mathcal{A}, \mathcal{E}] = |P_r[W_0] - P_r[W_1]|, \qquad (2)$$

where $P_r[W_b]$ is the probability that $\mathcal{A}$ calculates 1 given the challenger has selected b at the beginning of the game.

Now we define a pseudo random number generator attack game. Given a pseudo random number generator $\mathcal{G}$ defined over $(\mathcal{S}, \mathcal{R})$. The game starts with the challenger randomly selecting a binary value $b \in \{0,1\}$. If $b = 0$ the challenger generates a random seed $s \xleftarrow{R} \mathcal{S}$ and bit stream $\mathcal{G}(s) \in \mathcal{R}$ and sends it to the adversary $\mathcal{A}$. Otherwise if $b = 1$ then the challenger creates a truly random bit stream $r \xleftarrow{R} \mathcal{R}$ and sends it to the adversary $\mathcal{A}$. Finally the adversary computes the value of b. $\mathcal{A}$'s advantage with respect to $\mathcal{G}$ is

$$PRGadv[\mathcal{A}, \mathcal{G}] = |P_r[W_0] - P_r[W_1]|. \qquad (3)$$

To evaluate the total system security we must understand the most likely attacks. There are at least four primary attacks on this system. The first is replacement of the sensor with a bogus device, the second is the replacement of the ASIC with a bogus device, the third is a passive eavesdropping (EA) attack between the sensor and ASIC and the last is the active man in the middle attack (MITMA) between the sensor and ASIC. Note that a passive attack can only read the transmitted data between the sensor and the ASIC, while an active attack can both read and write the transmitted data.

Security against sensor replacement is based on the difficulty of forging a valid certificate (from a trusted source) with an associated secret key. This is determined by the security of the public key signing algorithm used for the certificate such at DSA [4] or ECDSA [5] and the security of the secret key associated with the certificate (stored in ROM).

There is no cryptographic security against replacement of the ASIC by an attacker in this system. The sensor will send data to any receiver that can negotiate a valid connection. Making it physically difficult to replace the ASIC is the only level of security.

There is no cryptographic security against EA in this system. Therefore, image sensor data can be freely collected by an attacker and used for any nefarious purpose.

Security against data modification by a MITMA depends not only on the MAC security, but also the security of the secret key. The security of the secret key is function of the random number generator in the ASIC, the security of the asymmetric encryption algorithm and side channel based key recovery security of the sensor and the ASIC. Using the Union Bound for probabilities the MITMA security can be bounded by

$$MITMAadv[\mathcal{A}, \mathcal{I}, \mathcal{E}_{pk}, \mathcal{G}, \mathbb{S}, \mathbb{C}] \leq \qquad (4)$$
$$MACadv[\mathcal{B}_{\mathcal{I}}, \mathcal{I}] + CCAadv[\mathcal{B}_{pk}, \mathcal{E}_{pk}] +$$
$$PRGadv[\mathcal{B}_{\mathcal{G}}, \mathcal{G}] + SCAadv[\mathcal{B}_{\mathbb{S}}, \mathbb{S}] +$$
$$SCAadv[\mathcal{B}_{\mathbb{C}}, \mathbb{C}],$$

where $MACadv[\mathcal{B}_{\mathcal{I}}, \mathcal{I}]$ is the advantage of the MAC $\mathcal{I}$. Examples of $\mathcal{I}$ include HMAC [6], CMAC [7] or GMAC [8]. $CCAadv[\mathcal{B}_{pk}, \mathcal{E}_{pk}]$ is the chosen cipher text advantage of the public key encryption algorithm $\mathcal{E}_{pk}$. Examples of $\mathcal{E}_{pk}$ include RSA [9] or ECC [10]. $PRGadv[\mathcal{B}_{\mathcal{G}}, \mathcal{G}]$ is the advantage of the pseudo random number generator $\mathcal{G}$. Examples of $\mathcal{G}$ include Salsa20 [11], ChaCha20 [12] or a true random number generator [13]. $SCAadv[\mathcal{B}_{\mathbb{S}}, \mathbb{S}]$ and $SCAadv[\mathcal{B}_{\mathbb{C}}, \mathbb{C}]$ are the advantages of side channel based key recovery attacks against the image sensor $\mathbb{S}$ and ASIC $\mathbb{C}$ respectively. Each $\mathcal{B}_x$ is a sub adversary of $\mathcal{A}$. Therefore this system can only be secure against MITMA if all of the pixels in each image are MACed, the MAC algorithm is secure, the public key encryption algorithm is secure, the random number generator is secure and the image sensor and ASIC are secure against side channel attacks.

Penetration testing is a critical part of cyber security system design. This process enables designers to empirically determine $SCAadv[\mathcal{B}_{\mathbb{S}}, \mathbb{S}]$ and $SCAadv[\mathcal{B}_{\mathbb{C}}, \mathbb{C}]$ as functions of the amount of data encrypted. Therefore, bounding how often the secret keys must be updated to achieve a desired level of security.
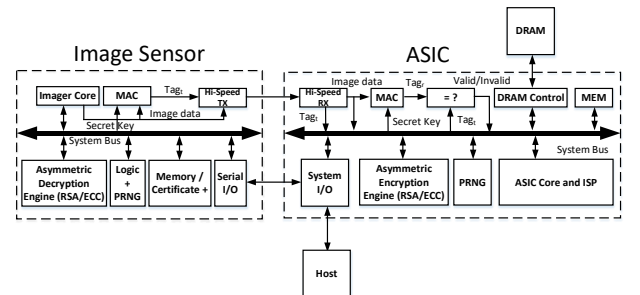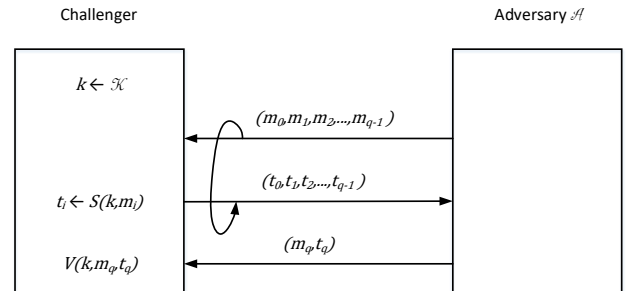

Figure 1

Figure 2

III. Next Generation Cyber Security

Many of the drawbacks of the current cyber security system described in Section II can be corrected using the system shown in Figure 3. In Figure 3 the image sensor is connected to the ASIC in the same manner as the previous system, but it incorporates symmetric certificate exchange between the sensor and the ASIC. In addition, authenticated encryption (AE) [14] is used for all of the image sensor data and the command data between the image sensor and the ASIC. This enables both security and integrity of the data to and from the image sensor. The improved boot-up sequence between the image sensor and the ASIC starts by having the ASIC and image sensor both exchange certificates. After both of the signatures of the certificates are validated, using global public keys, the public keys from each certificate are loaded into the respective asymmetric encryption hardware. Next the image sensor and the ASIC randomly create secret key materials for the authenticated encryption blocks which are then encrypted using the respective public keys from the exchanged certificates. The encrypted secret key materials are shared between the image sensor and the ASIC. Then both the image sensor and the ASIC combine the key materials to create the final secret key(s). This can be done using a collision resistant hash function [15], such as SHA-256 [16] or using an elliptic curve multiplication depending on the algorithm used to exchange the key materials. Finally the hashed secret keys are loaded into the AE hardware and the encryption process for the image data and the command stream begins. Note that the secret keys used for image data and command data must be separate (therefore we need 4 separaete secret keys).

The proposed system in Figure 3 is not without limitations. First it requires significantly more processing logic and power than the system described in Section II. In addition it also makes command communication between the sensor and the ASIC much more complex. Usually commands between a sensor and an ASIC are a few bytes, but using AE for command data security makes the smallest package size about 48 bytes for a 16 byte or less command. This is because AE algorithms require that each transmission include a NONCE, cyber text and a tag. Typically the NONCE, cyber text and tag are at least 128 bits. The longer the command the lower the overhead, but this is a significant cost for security.

Just like in Section II there are four primary attacks on this system. The first two are replacement of the sensor or the ASIC by an attacker. Security against sensor or ASIC replacement is based on the difficulty of forging a valid certificate (from a trusted source) with an associated secret key. This is determined by the security of the public key signing algorithm used for the certificate and the security of the secret key associated with the certificate.

Security against EA in this system is based on the security of the symmetric encryption algorithm used as a part of AE and the ability of the system to keep the secret keys safe. Since this system uses bi-directional certificates and both the image sensor and the ASIC create parts of the secret keys, the probability of breaking the cipher text from the sensor and from the ASIC is very low. In addition, since both the image

sensor and the ASIC create part of the key materials, even if the entropy of the PRGs is low such as ½ bit per bit, after the key materials are combined (using a cryptographic hash like SHA-256) the total entropy of the final secret keys should be very close to 1 bit per bit. Again using the Union Bound we find the EA advantage of the system

$$EAadv[\mathcal{A}, \mathcal{E}_{sk}, \mathcal{E}_{pk}, \mathcal{G}, \mathbb{S}, \mathbb{C}] \leq \qquad (5)$$
$$CCAadv[\mathcal{B}_{\mathcal{E}_{sk}}, \mathcal{E}_{sk}] + CCAadv\left[\mathcal{B}_{\mathcal{E}_{pk}}, \mathcal{E}_{pk}\right]^2 +$$
$$PRGadv[\mathcal{B}_{\mathcal{G}}, \mathcal{G}]^2 + SCAadv[\mathcal{B}_{\mathbb{S}}, \mathbb{S}] +$$
$$SCAadv[\mathcal{B}_{\mathbb{C}}, \mathbb{C}],$$

where $CCAadv[\mathcal{B}_{\mathcal{E}_{sk}}, \mathcal{E}_{sk}]$ is the advantage of the symmetric encryption algorithm used in the AE, such as GCM [8].

Just as in Section II security against MITMA depends not only on the MAC security of the AE algorithm, but also the security of the secret keys. The security of the secret keys is a function of the random number generator in the sensor and the ASIC, the security of the asymmetric encryption algorithm and side channel attack security of the sensor and the ASIC. Using the Union Bound for probabilities the MITMA security can be bounded by

$$MITMAadv[\mathcal{A}, \mathcal{I}, \mathcal{E}_{pk}, \mathcal{G}, \mathbb{S}, \mathbb{C}] \leq \qquad (6)$$
$$MACadv[\mathcal{B}_{\mathcal{I}}, \mathcal{I}] + CCAadv\left[\mathcal{B}_{\mathcal{E}_{pk}}, \mathcal{E}_{pk}\right]^2 +$$
$$PRGadv[\mathcal{B}_{\mathcal{G}}, \mathcal{G}]^2 + SCAadv[\mathcal{B}_{\mathbb{S}}, \mathbb{S}] +$$
$$SCAadv[\mathcal{B}_{\mathbb{C}}, \mathbb{C}].$$
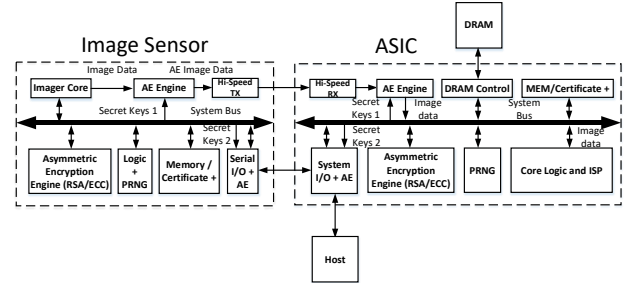

Figure 3

IV. Discussion

Bidirectional authentication and key exchange help to significantly improve the security of the proposed system. First bidirectional certificate exchange validates the identity of the components on both sides of the communication link. Then it reduces the security requirements for both the key generation process in the sensor and in the ASIC. It also reduces the security requirements for the asymmetrically encrypted key materials sent between the devices. This is true because an attacker needs both the sensor and the ASIC key materials to recover the final secret key(s) used for AE.

Since encryption, message authentication and digital signatures are brittle to even a single bit error, data transmission errors cannot be distinguished from cyber-attacks. For example, an automotive image sensor can generate 10^{13} bits/hour, but if the bit error rate in a given video

transmission channel is on the order of 10^{-12} then there will be multiple corrupted frames per hour. In order to mitigate this problem, some type of error correction code is required in the final system for at least the video data (note that the command data rate is much lower than the video data). For example a (18,16) Solomon Reed [17] code, i.e. a code with 18 bytes per block including 2 parity bytes, can correct up to one byte per block and detect up to two byte errors per block. If the data transmission error rate is 10^{-12}, and the errors are assumed to be independent, then the probability of having a block that has uncorrected errors, assuming a (18,16) Solomon Reed coded channel, is

$$P_r(> 0 \text{ bit errors in a byte}) \tag{7}$$
$$= 1 - (1 - 10^{-12})^8 = p'$$

$$P_r(> 1 \text{ bytes in a block have bit errors}) \tag{8}$$
$$= 1 - \sum_{i=0}^{1} \binom{18}{i}(1 - p')^{18-i}(p')^i$$
$$= 2 * 10^{-20}.$$

This would increase the expected time between uncorrected errors to

$$\frac{128}{(10^{13} * 2 * 10^{-20}) * 24 * 365.25} \tag{9}$$
$$> 73K \text{ years}.$$

Using error correction codes clearly reduces the error rate to an acceptable level, but it also increases the computation, power dissipation and chip size.

Cyber security is always a tradeoff between computation and performance. This means that enhanced security increases power dissipation, silicon area and system cost. Therefore, understanding the key attack scenarios, attack consequences, and mitigations is critical for optimizing the system for a required security level.

V. CONCLUSIONS

We have shown that current cyber security systems in image sensors can be insecure under certain conditions. These conditions include, eavesdropping attacks between the source and destination, not MACing all of the data in a message, control of the sensor or ASIC and key recovery attacks. We have proposed a next generation cyber security system that tries to mitigate most of the current generation's problems using symmetric certificate exchange and symmetric secret key material exchange in addition to adding authenticated encryption to both the image and command data channels.

Although the proposed cyber security architecture is more secure than current systems, it is still sensitive to key recovery attacks especially against the static private keys associated with the certificates. To further improve security an online certificate status protocol (OCSP) [18] interface could be implemented in the system to check the validity of all of the component certificates. If a given component certificate is invalid then data from that component would also be considered invalid in the system. Finally the proposed cyber security architecture does not include error correction which is necessary to detect active attacks on the system.

REFERENCES

[1] D. Boneh and V. Shoup, "A Graduate Course in Applied Cryptography", http://crypto.stanford.edu/~dabo/cryptobook/bonehshoup_0_4.pdf.

[2] R. Spreitzer et al., "Systematic Classification of Side-Channel Attacks: A Case Study for Mobile Devices", IEEE Communications Surveys & Tutorials, Vol. 20, No. 1, 2018.

[3] https://en.wikipedia.org/wiki/X.509.

[4] M. Bellare et al., "The Exact Security of Digital Signatures-How to Signwith RSA in Rabin", Advances in Crytology – EUROCRYPT '96, EUROCRYPT 1996, Lecture Note in Computer Science, vol. 1070, Springer, Berlin, Heidelberg.

[5] D. Johnson et al., "The Elliptic Curve Digital Signature Algorithm (ECDSA), IJIC 1, 36-63 (2001).

[6] P. Gauravaram, S. Hirose and S. Annadurai, "An Update on the Analysis and Design of NMAC and HMAC Functions", International Journal of Network Security, Vol.7, No.1, PP.49–60, July 2008.

[7] C. Baritel-Ruet, F. Dupressoir, P. Fouque and B. Gregoire, "Formal Security Proof of CMAC and its Variants", 2018 IEEE 31st Computer Security Foundations Symposium.

[8] A. Delignat-Lavaud et al., "Implementing and Proving the TLS 1.3 Record Layer", 2017 IEEE Symposium on Security and Privacy.

[9] R.L. Rivest, A. Shamir, and L. Adleman, A method for obtaining digital signatures and public-key cryptosystems, Commun. ACM, Feb. 1978, 21(2): 120-126.

[10] V. Kapoor et al., "Elliptic Curve Cryptography", ACM Ubiquity, Volume 9, Issue 20, May 20 – 26, 2008.

[11] D. J. Bernstein, "The salsa20 family of stream ciphers," eSTREAM, ECRYPT Stream Cipher Project, Report 2005/025, 2005, http://www.ecrypt.eu.org/stream.

[12] D.J. Bernstein, "Chacha, a variant of salsa20," Jan. 2008, http://cr.yp.to/chacha.html.

[13] B. Sunar et al., "A Provably Secure True Random Number Generator with Built-In Tolerance to Active Attacks", IEEE Transactions on Computers, vol. 56, no. 1, January 2007

[14] M. Bellare et al., "Authenticated encryption in SSH: Provably fixing the SSH binary packet protocol", ACM Conference on Computer and Communications Security (CCS-9) (2002), ACM Press, pp. 1–11.

[15] https://en.wikipedia.org/wiki/Collision_resistance

[16] NIST/NSA, "FIPS 180-2: Secure Hash Standard (SHS)", August 2002 (change notice: February 2004).

[17] S. B. Wicker and V. K. Bhargava, "Reed-Solomon codes and their applications", John Wiley & Sons, 1999.

[18] https://en.wikipedia.org/wiki/Online_Certificate_Status_Protocol

A CMOS Image Sensor With 1.6us Conversion Time 10-bits Column-Parallel Hybrid ADC Using Self-Adaptive Charge-Injection Cell

Min Ruei Wu, Yu-Hsiang Huang and Chih-Cheng Hsieh, *Senior Member, IEEE*

Abstract— **This paper presents a 256x256 CIS using the proposed column-parallel hybrid SA-CI-SS ADCs fabricated in TSMC 40nm. The column ADC consumes a total power of 25.6uW@400KS/s at 1.5/1V (SH&Comp / digital&CI) operation. The achieved ENOB is 9.41-bit with a DNL/INL of -0.32/0.49LSB and -0.37/0.62LSB, respectively. The achieved nonuniformity of 256 column ADCs is within 0.22% (standard deviation) and 1.08% (peak-to-peak) with a test input range of 0.8V~1.7V. With a pitch of 4um, this work achieves a 10-bit conversion in 1.6us@25.6uW at a 40Mhz clocking clock and a state-of-the-art FoM of 0.25 um*fJ/c.-s.**

I. INTRODUCTION

In recent years, with the Internet of Things (IoT) being widely used in our daily life, the demand of CMOS image sensors (CIS) with high frame rate has also increased rapidly. Therefore, the conversion speed of the analog-to-digital converter (ADC) in the readout circuit of CIS is gradually becoming the bottleneck. The column-parallel single-slope (SS) ADCs [1] are widely adopted due to its small area for column-pitch implementation and high energy efficiency, but the operation speed decreases exponentially with an increased resolution. For high-resolution high frame-rate imaging applications, the required clock frequency is up to GHz range, which causes the burdens of clock distribution and power consumption. To omit the required high-frequency counting clock of SS ADC, several works have been reported [2-4]. The time-stretched (TS) SS ADC [2] implemented a two-step conversion with time residue expansion. However, the required V-T-V converter needs two large capacitors for expansion ratio implementation and thermal noise suppression with the area and power penalty. The time-to-digital converter (TDC) interpolation SS ADCs [3] implemented a delay-chain-generated multiple phase clock for TDC operation, but still needs the high-frequency counting and complex delay calibration. A capacitor array-assisted charge-injection (CI) SAR ADC [4] was reported using coarse-fine CI-arrays for conversion step and energy reduction, however, a complex calibration is required to satisfy the required matching and linearity performance. Moreover, the unit charge of Vth-based CI-cell is sensitive to PVT variation and needs extra reference voltages for CI.

II. PROPOSED COLUMN-PARALLEL HYBRID ADC

To address the mentioned issues, this paper presents a column-parallel hybrid ADC using self-adaptive (SA) CI-cell and SS conversion. Compared to the pure SS ADC, the proposed hybrid architecture effectively reduces the conversion cycles (for 10-bit resolution) from 1024 to 64 with a coarse-fine operation, which achieves a significant power reduction by using a 10x slower counting clock of 40Mhz. Compared to the reported CI SAR ADC, the proposed hybrid architecture using SA CI-cell achieves a high conversion linearity in a small area without need of CDAC and weighting calibration. Compared to the conventional Vth-based CI-cell using MOSFET threshold voltage (Vth) [4] for unit-charge control, which suffers from the PVT variation and slow settling, the proposed SA CI-cell achieves a high-speed constant unit-charge injection amount using a feedback circuit and self-adaptive operation. By applying a ramping reference on portion of sampling capacitance for residue's fine SS conversion, the coarse-fine weighting is guaranteed by a local capacitance matching without calibration. To reduce the power consumption and area further, a global double-data-rate (DDR) 6-bit gray code counter is also implemented for SS operation.

A. Chip architecture Overview

Fig. 1 shows the chip architecture of the prototyped CIS and the block and critical timing diagrams of the proposed column-parallel hybrid CI/SS ADC. The column-ADC consists of the sample/hold circuit, SA CI-cell, dual-mode comparator, and CI/SS SRAM for data storage. In the coarse asynchronous CI phase, after the pixel signal (V_{sig}) sampling on the top plate (V_{top}) of sampling capacitor C_T ($=C_L + C_R$), the Vtop is steply ramped down by periodically transferring a constant unit charge Q_{ci} from C_T to C_{ci}. When $V_{top} < V_{ref}$, the comparator output (Valid) triggers and latches the counting code on CI SRAM from the global gray counter operating at a 40MHz counting clock. Then, in
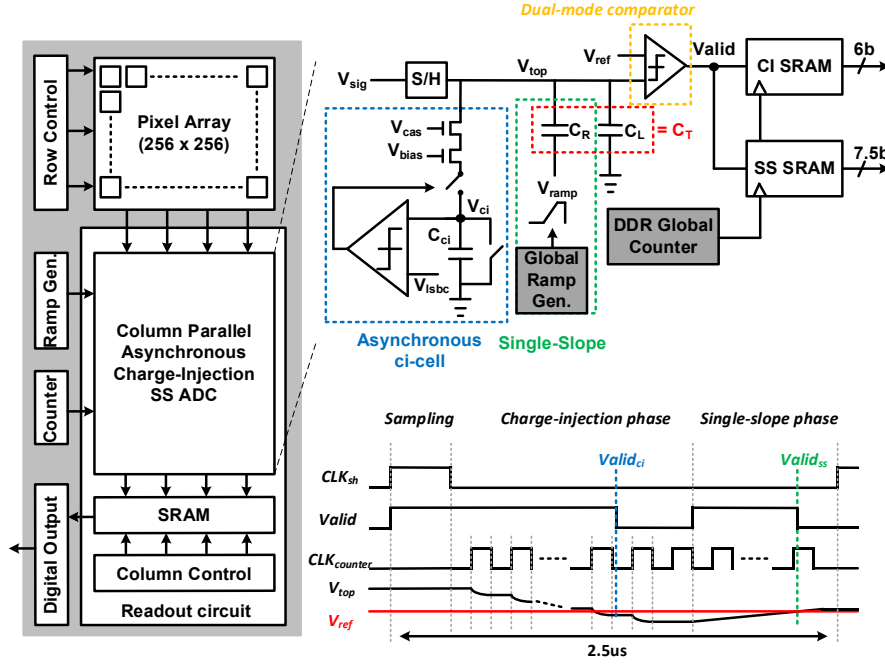
Fig. 1. CIS prototype architecture and the proposed column-parallel hybrid CI/SS ADC.

the fine SS conversion phase, the residue on V_{top} is ramped up by applying a V_{ramp} on the bottom plate of C_R from the global ramp generator. Similarly, when $V_{top} > V_{ref}$, the comparator output triggers and latches the counting code on SS SRAM. With the proposed CI/SS hybrid architecture, the prototyped ADC can finish a 10-bits conversion in 1.6us by reusing a 6-bits DDR gray counter operated at 40MHz.

B. The Proposed SA CI operation

Fig. 2 shows the comparison of the V_{th}-based and proposed SA CI operations. At the beginning with T_{pre}= high, CT is pre-charged to V_{ref}. For each toggling period of "EN" and "inject", a unit charge on CT is transferred to Cci by "EN=1" and then discarded by "inject=1". In the conventional V_{th}-based CI operation, the unit charge is defined as $Q_{ci.Vth} = C_{ci}*\Delta V_{ci}$, where $\Delta V_{ci.Vth} = V_{bias} - V_{th.Mci}$ and is PVT-sensitive and time-variant from the subthreshold settling behavior. The settling accuracy requirement limits the achievable conversion speed, and a complex calibration is necessary for PVT variation. To address the mentioned issues, this work proposes a self-adaptive asynchronous CI-cell using comparator and feedback mechanism to realize a constant unit charge injection amount. The achieved voltage difference $\Delta V_{ci.SA} = V_{lsbc}$ and the corresponding unit charge $Q_{ci.SA}$ are PVT-insensitive, time-invariant, and well-controlled by the closed-loop operation.

The operation of SA CI is explained in detail as follows. During the reset phase with CLK_{ci} = low, the switch SW1 turns off and SW2 turns on to reset $V_{ci.SA}$ = 0. Simultaneously, an auto-zeroing operation (triggered by CLK_{az}) with C_{az} is applied on the
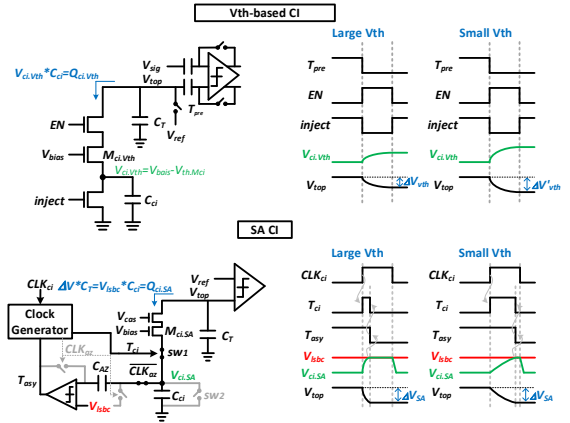
Fig. 2. Comparison of the conventional Vth-based and the proposed SA CI operations.

comparator to eliminate the static offset and reduce the column-to-column gain mismatches. Then in CI phase with CLK_{ci} = high, SW2 turns off and SW1 turns on, the signal charge on C_T is transferred to C_{ci} and increase $V_{ci.SA}$. When $V_{ci.SA} > V_{lsbc}$, the comparator output Tasy flips to pull down T_{ci} and turn off SW1 to finish the unit CI operation. By sensing the voltage difference $V_{ci.SA}$ using comparator instead of V_{th} of MOSFET, the CI operation is asynchronous and self-adaptive to provide a constant unit charge injection amount $Q_{ci.SA} = C_{ci} * V_{lsbc}$. The 5-bits coarse conversion is achieved by repeating the CI operation (Reset and CI phase) 32 times (at most) without consuming any conversion energy by utilizing the signal charge sampled on C_T, and without need of reference voltage. The cascode structure biased by V_{cas} and V_{bias} is implemented to guarantee a constant V_{ds} of

M_{ci} to reduce the $Q_{ci.SA}$ error from the signal-dependent parasitic capacitance of SW1 and improve linearity. The LSB weighting depends on the ratio of C_{ci}/C_T and V_{lsbc}/V_{swing} is configurable, where V_{swing} is the full swing of ADC input range.

C. Fine SS conversion implementation

Fig. 3 shows the fine SS conversion implementation with dual-mode comparator. After the coarse SA CI conversion, the residue on V_{top} is ramped up by applying a global V_{ramp} on C_R. The coarse-fine weighting is guaranteed by the local capacitance ratio matching of C_{ci}/C_R and the global reference voltage ratio of V_{lsbc}/V_{ramp}, which is immune to the column-to-column mismatch of sampling capacitor C_T and achieves a good uniformity without the need of calibration. For a better energy efficiency, a dual-mode comparator is implemented for the dynamic and static comparison operations of SA CI and SS conversions, respectively. Since the crossing point of comparisons in coarse and fine conversions are all at V_{ref}, there is no dynamic offset error. To cover the offset mismatch between dynamic and static comparison operations, 5-bit coarse and 6-bit fine conversions are implemented to get the 10-bit result with a 1-bit redundancy. The column fixed-pattern-noise (CFPN) from the column-to-column offset mismatch of dual-mode comparator can be easily cancelled out using a dark reference calibration.

III. Measurement Results and Conclusion

A 256x256 CIS using the proposed column-parallel hybrid SA-CI-SS ADCs is prototyped in TSMC 40nm. The column ADC consumes a total power of 25.6uW@400KS/s at 1.5/1V (SH&Comp / digital&CI) operation. Fig. 4 shows the measured dynamic and static performance. The achieved ENOB is 9.41-bit with a DNL/INL of -0.32/0.49LSB and -0.37/0.62LSB, respectively. Figure 5 shows the ADC array uniformity performance. The achieved nonuniformity of 256 column ADCs is within 0.22% (standard deviation) and 1.08% (peak-to-peak) with a test input range of 0.8V~1.7V. Figure 6 shows the captured images before and after dark calibration. It shows the CFPN of captured image is successfully cancelled out by a dark frame calibration. Table I shows the comparison table with the state-of-the-art works [1-4]. With a pitch of 4um, this work achieves a 10-bit conversion in 1.6us@25.6uW at a 40Mhz clocking clock and a state-of-the-art FoM of 0.25 um*fJ/c.-s.

Acknowledgment
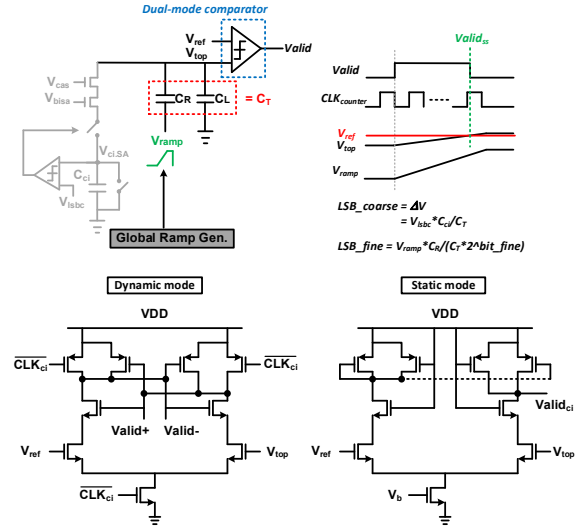
Fig. 3. Fine SS conversion using dual-mode comparator.
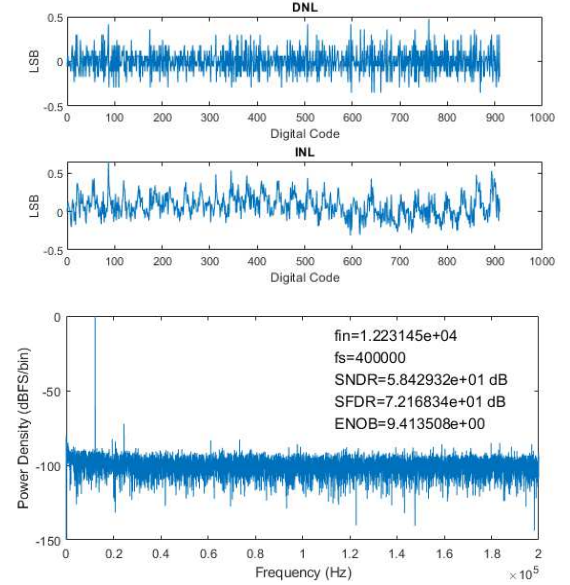
Fig. 4. Measurement results.
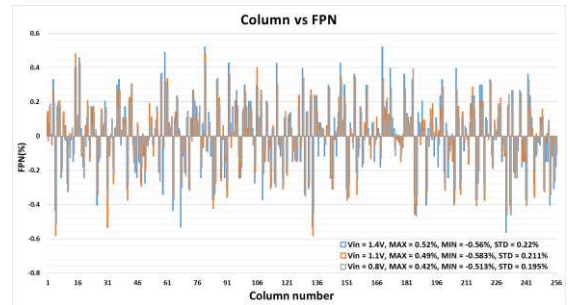
Fig. 5. The measured uniformity of ADC array with an input of 0.8~1.4V

also acknowledge the support of Taiwan Semiconductor Research Institute (TSRI) and Taiwan Semiconductor Manufacturing Company (TSMC) for the fabrication of the test chip.

REFERENCES

[1] T. Toyama, et al., "A 17.7Mpixel 120fps CMOS image sensor with 34.8Gb/s readout," 2011 ISSCC, 2011, pp. 420-422.
[2] I. Park, et al., "A 76mW 500fps VGA CMOS Image Sensor with Time-Stretched Single-Slope ADCs Achieving 1.95e-Random Noise," ISSCC, 2019, pp. 100-102.
[3] D. Levski, et al., "A 1- us Ramp Time 12-bit Column-Parallel Flash TDC-Interpolated Single-Slope ADC With Digital Delay-Element Calibration," TCAS I, 2019, pp. 54-67.
[4] K. D. Choo, et al., "Energy-Efficient Motion-Triggered IoT CMOS Image Sensor With Capacitor Array-Assisted Charge-Injection SAR ADC," JSSC, 2019, pp. 2921-2931.
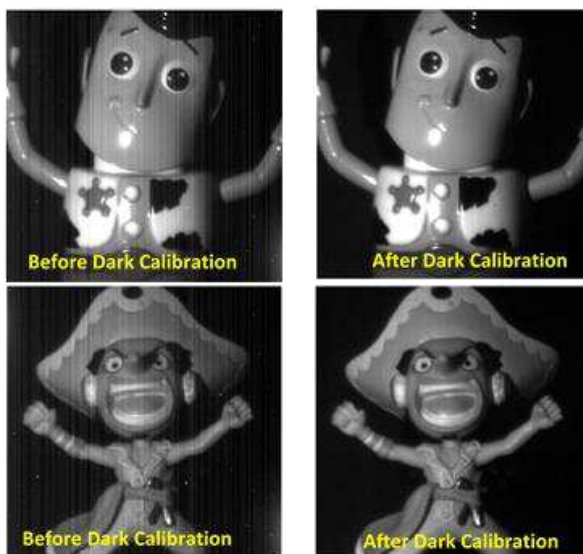
Fig. 6. Captured images before/after dark calibration
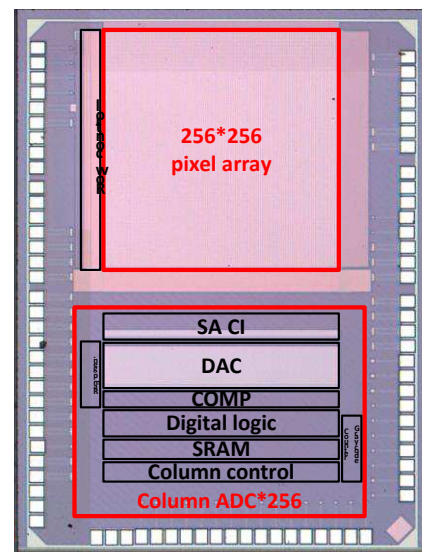
Fig. 7. Chip micrograph.

TABLE I
COMPARISON TABLE WITH STATE-OF-THE-ART WORKS.

	ISSCC 2011 [1]	ISSCC 2019 [2]	TCAS I 2019 [3]	JSSC 2019 [4]	This work
Technology	90nm	110nm	130nm	65nm	40nm
ADC technique	SS	SS/Time-stretched	SS/TDC	SAR/CI	CI/SS
Pixel array	8192*2160	640*480	1024*128	792*528	256*256
Frame-rate (fps)	120	500*	1776*	5.6	1563
1-H time (us)	3.9	4	4.4	370	2.5
Pixel supply (V)	2.9	3.3	3.3	1.7	2.5
ADC supply (V)	2.7/1.2	3.3/1.5	3.3/1.5	1.7	1.5/1
ADC input range (V)	-	0.8	1.5-2.9	0.5	1
Counter clk (MHz)	2376	100	250	-	40
Resolution (bits)	12	10	12	10	10
ADC pitch (um)	4.2	8	5.6	3.1	4
ADC power (uW)/Col	366	98.125	177	0.163*	25.6
FoM (um×pJ/c.-s.)	1.463	3.066	1.064	0.1825	0.25

* Calculated $\text{FoM} = \dfrac{\text{per column ADC power(uW)} \times \text{ADC pitch(um)} \times 1-\text{H time(us)}}{2^{\text{Resolution}}}$

Charge Demultiplexing for an Ultra-High-Speed Charge-Domain CMOS TDI Image Sensor with a multi-MHz Line Rate

Hyun Jung Lee, Suzy Patchett, David Atos, Nixon O, and Paul Donegan

Teledyne DALSA, Teledyne Digital Imaging, Inc.
605 McMurray Rd., Waterloo, ON, Canada N2V 2E9
Tel: +1-519-886-6000 / E-mail: hyunjung.lee@teledyne.com

Abstract— **Charge demultiplexing for an ultra-high-speed charge-domain CMOS time delay and integration (TDI) image sensor is presented. A proposed charge steering structure enables charges transferred through multiple TDI rows in a column to be demultiplexed to the corresponding sub-column sense nodes (SNs) for parallel processing. After charge demultiplexing, the signal charges in the sub-column SNs are converted in column readout circuits in parallel. This conversion operation is also performed simultaneously while charges in the next TDI rows are transferred and demultiplexed. In contrast to the slow, sequential operation of a conventional CMOS TDI imager, we show that our proprietary charge demultiplexing for parallel processing is a disruptive technology that can increase a CMOS TDI scan speed to an unprecedented level (i.e., multi-MHz) for highly demanding applications.**

I. INTRODUCTION

After the successful demonstration of charge-domain TDI imaging using a conventional CMOS technology [1], we have been rapidly displacing our traditional CCD TDI image sensors with CMOS counterparts in various applications such as industrial machine vision [2], DNA sequencing and Earth observation [3]. Especially for Earth observation, our imager is the first space-grade charge-domain CMOS TDI imager launched into orbit in the world to the best of our knowledge. One of the remaining applications that is not accessed with CMOS TDI yet is pattern defect inspection on semiconductor masks and wafers [4], in which ultra-high-speed (i.e., a MHz TDI scan rate) and ultra-high-sensitivity (i.e., a thousand TDI stages) imaging capabilities are essential.

In a conventional CMOS TDI imager, however, a pixel array is read out row by row. This sequential operation limits sensor speed. A maximum TDI scan rate in the current state-of-the-art imager is several hundred kHz [2]. In contrast to the traditional readout, higher speed operation can be achieved by processing multiple TDI rows in parallel. In this work, we present charge demultiplexing for parallel processing for an ultra-high-speed charge-domain CMOS TDI imager.

II. CHARGE DEMULTIPLEXING

Figure 1 shows the block diagram of a column slice of our proprietary charge demultiplexing CMOS TDI pixel array [5]. Here, a single column consists of the m number of active (denoted as white) TDI imaging pixels, out of which the n number of pixels is to be processed in parallel, and the corresponding n number of light-shielded (denoted as gray) sub-column output structures at the end of the pixel array. Not explicitly shown here, the output structure is located at both the top and bottom of the array for bi-directional scanning operation. Each of the sub-column output structures has a charge steering gate (CST) in addition to a typical pixel readout structure such as a SN, a reset, and a source follower.
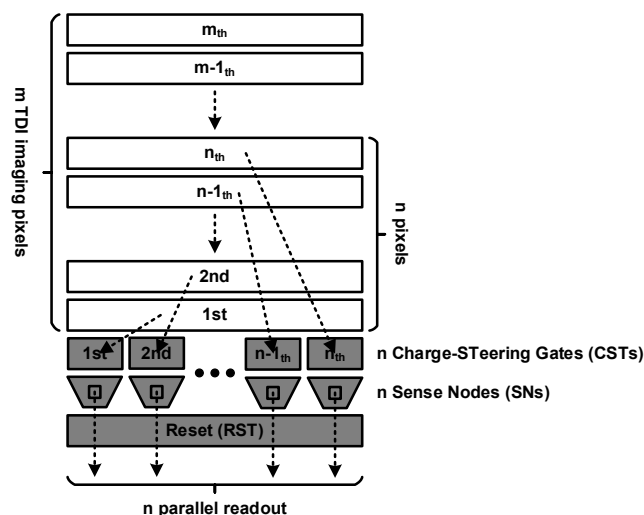
Figure 1. Block diagram of a column slice of the proposed charge demultiplexing TDI pixel array.
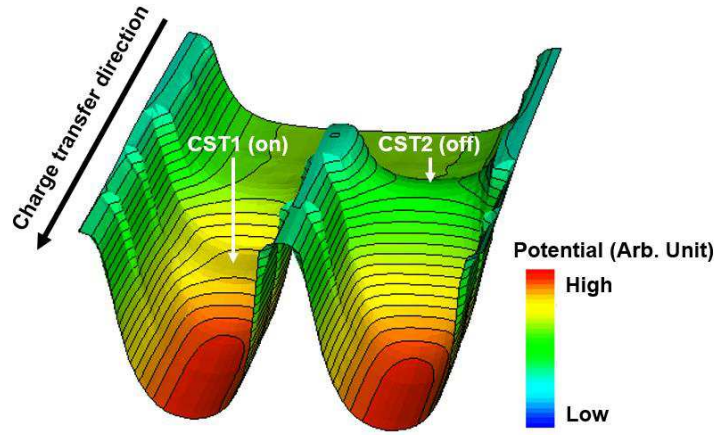
Figure 2. Surface plot of the computed electrostatic potential maximum for 2-row charge steering operation.

The proposed charge demultiplexing is achieved by an operation so-called 'charge steering' which enables each of charges transferred vertically through the n TDI rows in a column to be demultiplexed horizontally to each of the corresponding n sub-column SNs. For this operation, each of the n CSTs in a single row turns on in sequence. Therefore, the CST should be capable of transferring and blocking charge completely when it is ON and OFF, respectively. A surface plot of the computed electrostatic potential maximum of the charge steering structure for 2-row operation is presented in Figure 2. Here, charge is transferred through the CST1 clocked high while blocked by the CST2 clocked low due to a potential gradient and a barrier formed to transfer charge only through the CST1 along the charge transfer direction. As the CSTs are added in the light-shielded region and charge is not stored under it, they will not affect most of the imager performance parameters such as full well capacity, linearity and responsivity. However, the incomplete charge transfer (and also incomplete charge block) will manifest itself as row-to-row charge mixing which will degrade the along-track modular transfer function (MTF) in the TDI operation. This can be confirmed by charge transfer (and also block) efficiency measurements. The test results will be discussed in the next section.

After all charge demultiplexing operations are completed, the signal charges in the sub-column SNs are converted in the corresponding column readout circuits in parallel. This conversion operation is also performed simultaneously while charges accumulated in the next n number of TDI rows are transferred and demultiplexed. The operational timing diagram of 2-TDI row demultiplexing implemented in our test chip is provided in Figure 3(a) in comparison with the diagram of the traditional sequential timing shown in Figure 3(b). In both the diagrams, SHR and SHS are reset and signal sampling periods, respectively, and t1 is a line period. In Figure 3(a), t1 is also a row transfer time, t2 and t3 are transfer periods for the present 2 TDI rows and for the next 2 TDI rows, respectively, and t4 and t5 are AD conversion windows for the signals from the previous 2 TDI rows and from the present 2 TDI rows, respectively. In Figure 3(b), t2 is a row transfer time which is the same duration as t1 in Figure 3(a), which illustrates that the 1-line period is reduced by the proposed parallel processing.

(a)   (b)
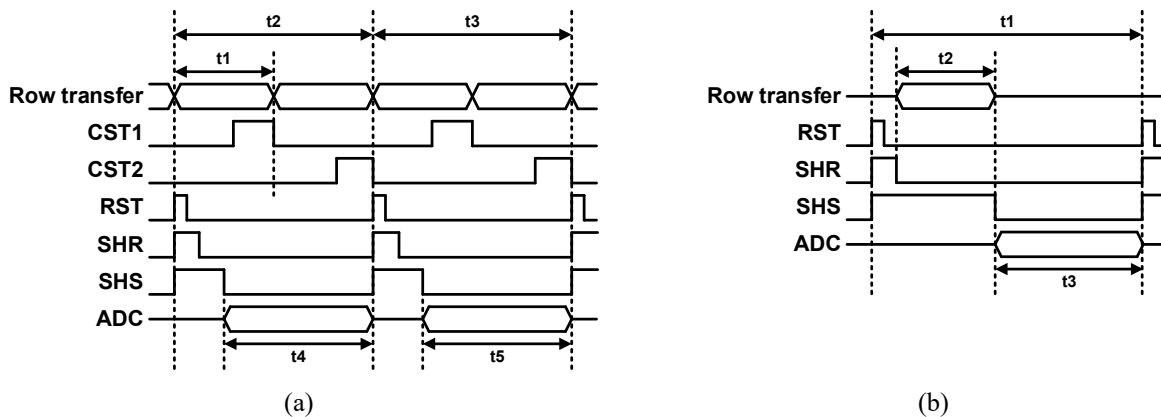
Figure 3. Operational timing diagram: (a) for the proposed parallel processing for 2 rows and (b) for the traditional sequential readout.

The charge demultiplexing operation can be extended to any higher number of rows to be processed in parallel. In real implementation, the 2-row steering can be cascaded repeatedly for better charge transfer. This is illustrated in Figure 6(a) for 4 (i.e., 2 × 2) rows as an example in comparison with direct charge steering without cascading in Figure 6(b). The signal charge (denoted as e-), for example, located at the rightmost corner of the first TDI row should traverse the entire row without cascading as shown in Figure 6(b), where fringing fields that aid charge transfer are essentially non-existent. However, it travels only half the distance in Figure 6(a).

(a)	(b)
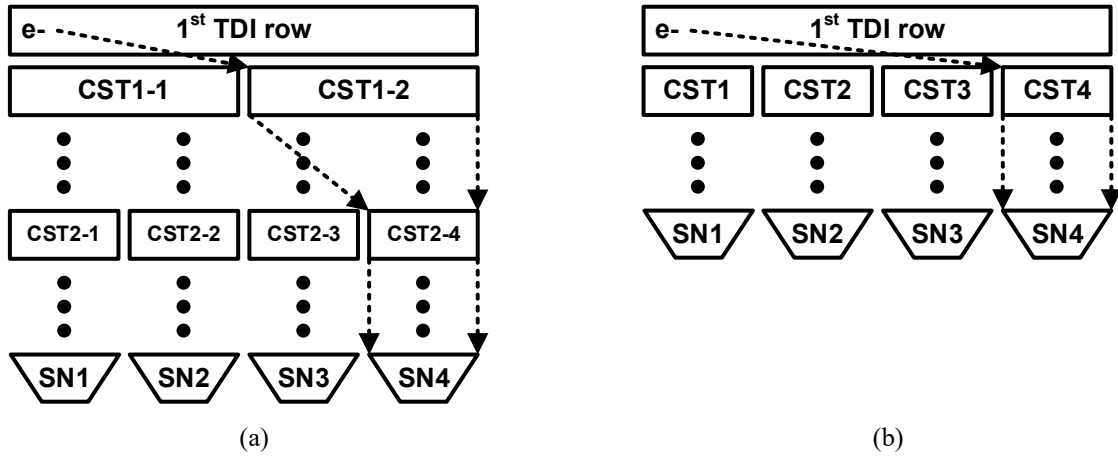
Figure 6. (a) Exemplary 2 × 2 cascaded charge steering operation and (b) steering without cascading.

The proposed charge demultiplexing increases the TDI scan rate greatly. The expected maximum line rate in 12 bit as a function of the number of parallel-processed TDI rows are calculated for two row transfer times, 200 ns and 400 ns, respectively in Figure 5. With 2-row (each row having a transfer time of 200 ns) parallel processing, a maximum TDI line rate is expected to be over 2 MHz. A line rate of 3 MHz was experimentally confirmed in 10 bits. The detailed test results are presented in the next section.
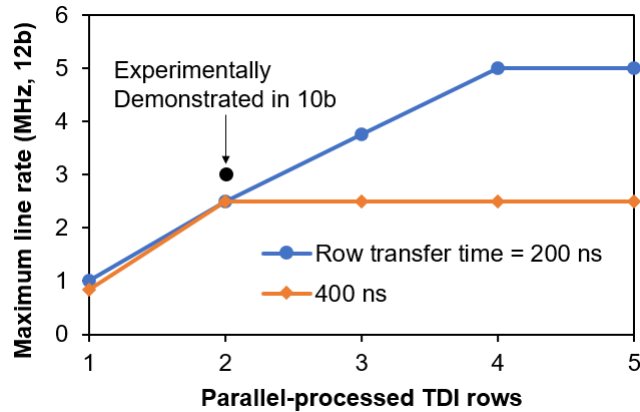
Figure 5. Expected maximum TDI scan rate as a function of the number of parallel-processed rows.

III. TEST RESULTS

A test pixel array including various charge demultiplexing structures for the 2-TDI row operation discussed earlier and CMOS peripheral circuits including high-speed pixel drivers and readout circuits were monolithically integrated in the test vehicle using the same 0.18 μm CMOS technology reported previously by our group [1]. Here, the 10-μm pitch active pixels were demultiplexed to two 5-μm pitch sub-columns. To assess the functionality of the proposed charge steering structure, charge transfer efficiency (CTE) was measured using the extended pixel edge response (EPER) method [6] at a transfer speed of 25 ns per transfer as shown in Figure 4, with the signal read out through the CST1 that is clocked high while the CST2 is clocked low. Here, the reduction in the signal of the first and last rows are not due to
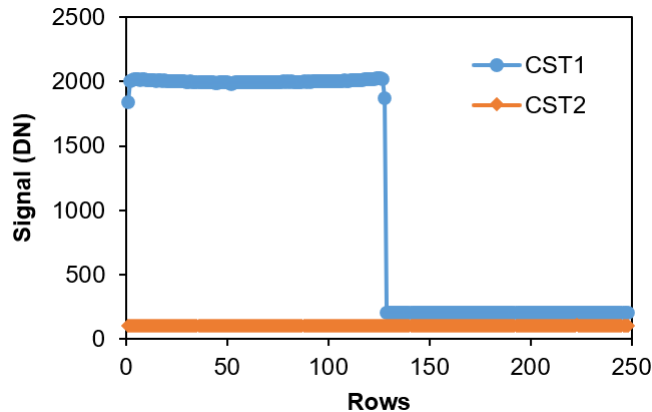
Figure 4. EPER for CTE measurement for the demonstration of charge demultiplexing.
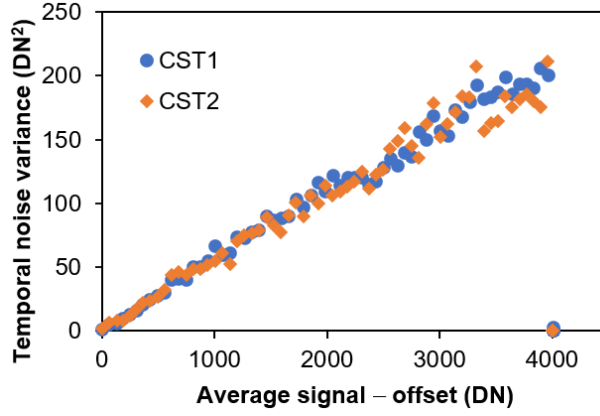
Figure 7. PTC.

incomplete charge transfer but due to the light-shield covering the first and last rows partly. The CTE over 0.99999 per transfer was measured through the CST1 while no charge leaked through the CST2 was detected. This demonstrates that the charge steering structure is capable of demultiplexing charge selectively to a wanted SN while blocking it to the other unwanted SN. The maximum TDI scan rate of 3 MHz was demonstrated experimentally in 10 bits with charge demultiplexing while the other imager performance parameters such as full well capacity, dark current, linearity and responsivity were not compromised. It is worth to note that the noise remains unchanged despite the higher line rate, since the speed improvement is achieved not by increasing the readout speed itself but by performing parallel processing at the same speed. The photon transfer curves (PTCs) for CST1 and 2 signals are presented in Figure 7. The test results are summarized in TABLE I.

TABLE I. Summary of the test results.

Parameters	Unit	Value
Pixel pitch	μm	10 (2×5-μm sub-columns)
Maximum TDI scan rate	MHz	3 (demonstrated in 10b)
CTE per transfer	Fraction of 1	> 0.99999
Conversion gain	μV/e-	13
Full well capacity	ke-	> 60
Dark current at 25°C	nA/cm^2	< 4
Non-linearity	%	< 2
Responsivity	DN/(nJ/cm^2)	4971
Noise floor	e-	~20

IV. CONCLUSION

We have developed charge demultiplexing for ultra-high-speed CMOS TDI imaging in charge domain. A test vehicle has successfully demonstrated a maximum TDI scan rate of 3 MHz without compromising the other imager performance. The ultra-high scan speed enabled by the proposed charge demultiplexing opens the door for highly demanding applications such as ultra-high-speed machine vision, semiconductor mask and wafer inspection, and flat panel display inspection. A charge demultiplexed CMOS TDI image sensor with high resolution is currently being developed, which will be reported in a future publication.

REFERENCES

[1] Hyun Jung Lee et al., "Charge-coupled CMOS TDI imager," International Image Sensor Workshop, 2017.
[2] https://www.teledynedalsa.com/en/products/imaging/cameras/linea-hs/.
[3] Owen Cherry et al., "Radiation-tolerant multispectral charge domain TDI CMOS imagers with integrated filters for Earth observation," Space and Scientific CMOS Image Sensors Workshop, 2019.
[4] Hiroki Miyai et al., "Actinic patterned mask defect inspection for EUV lithography," Proc. SPIE, vol.11148, 2019.
[5] https://patentscope.wipo.int/search/en/detail.jsf?docId=WO2022051835.
[6] James Janesick, "Scientific Charge-Coupled Devices," SPIE Press, 2001.

Detecting Short-wavelength Infrared Photons by Schottky-barrier based Single Photon Avalanche Diode in 180-nm CMOS Technology

Chun-Hsien Liu
Institute of Electronics
National Yang Ming Chiao Tung University
Hsinchu, Taiwan
terryliu225.ee07@nycu.edu.tw

Yu-Wei Lue
Institute of Electronics
National Yang Ming Chiao Tung University
Hsinchu, Taiwan

Sheng-Di Lin
Institute of Electronics
National Yang Ming Chiao Tung University
Hsinchu, Taiwan

Abstract—We propose a Schottky-barrier (SB) based single-photon avalanche diode (SPAD) in 180-nm CMOS technology. SB-based SPAD consisted of Schottky junction as active region can detect short-wavelength infrared photons by internal photoemission effect. The simulated strong electric field for triggering avalanche process ensures multiplication region locating at the active region. The preliminary measurement obtained the responsivity of ~60 mA/W in 1550 nm at 15 V, the dark count rate ranged 10 kHz – 2 MHz, and photon detection probability of ~0.35% at excess bias voltage of 1.0 V.

Keywords—Single photon avalanche diode, Short wavelength infrared sensor

I. INTRODUCTION

CMOS single-photon avalanche diode (SPAD) has excellent photo-sensitivity and timing response and serves as a receiver in light detection and ranging (LiDAR) [1]. For LiDAR application, short-wavelength infrared (SWIR) sensors have been attractive because of its safety level of maximum exposure intensity is much higher than that in near infrared regime and more laser power could give a longer detectable distance [2]. Due to silicon bandgap of 1.12 eV, silicon-based SPAD could not detect SWIR photons. To realize SWIR detection, researchers have turned to other materials such as Ge-on-Si SPAD, III-V SPAD or superconducting nanowire detector (SNSPD) [3-5]. In this work, we propose a Schottky barrier (SB) based SPAD fabricated in 180-nm CMOS technology with the detection wavelength range up to 1550 nm. The previous work [6] demonstrated SWIR photocurrent measurement using silicon-based Schottky diodes even the photon energy is lower than silicon bandgap due to the photon-absorption induced internal photoemission (IPE) effect [7]. Figure 1 illustrates the various SWIR photon detection mechanisms in a Schottky diode.

Path 1 is the best because the photo-generated energetic electrons have the highest efficiency to cross barrier. Path 2 is the cold-field emission and allows photon with energy lower than Schottky barrier height to be collected via quantum tunneling. The worst one is path 3 because the electrons could not cross the barrier to contribute any photocurrent. Therefore, the detectable longest wavelength is dominated by the Schottky barrier height. In this work, we propose a SWIR detector with Schottky-barrier based SPAD and show the simulate structure and breakdown simulation with technology computer aided design (TCAD) and descript our measurement environment in the section two. The device experiment results of electric and optical characteristics are shown and discussed in section three.

II. DEVICE AND MEASUREMENT

Figure 2 shows the designed structure of our Schottky-SPAD in 180-nm CMOS process without any customization. The device has a square-shape active area with a size of 10 μm. To detect SWIR, a Schottky barrier for photo-generated hot electrons to overcome from the metal into bulk silicon by IPE. The injected electrons are then accelerated by the underneath strong electric field and amplified as a voltage signal. The Schottky junction was formed between cobalt silicide (CoSi2) and high-voltage n-type well (HVNW) layer.

Figure 1. Photon-absorption induced internal photoemission (IPE) effect, and three possible paths, (1) hot electrons emission, (2) cold-field emission, and (3) recombination.
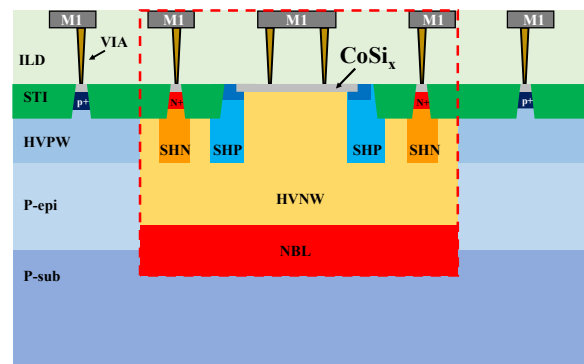
Figure 2. Schematic of the Schottky SPAD in cross-section view.
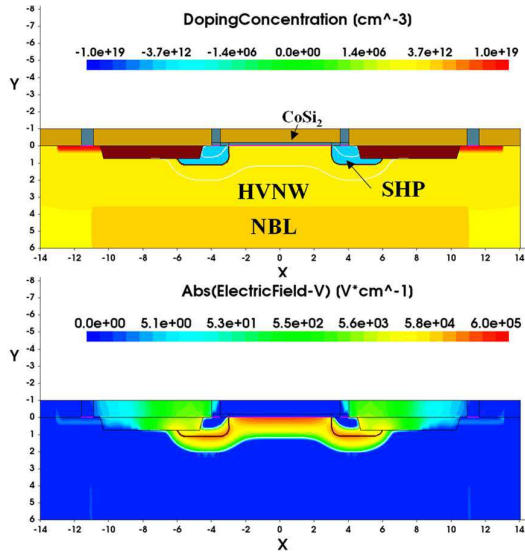
Figure 3. Simulated distributions of doping concentration (upper panel) and electric filed (lower panel). Simulation region is limited in the red dashed box in Fig 2.

The upper panel in Fig. 3 gives the distribution of doping concentration using TCAD simulation. For the lower panel in Fig. 3, we obviously observed the avalanche breakdown occurring at the Schottky junction. The maximum electric field is up to 70k V/cm, and impact ionization distribution spreads from y = 0 μm on surface to y = 1 μm. Figure 4 plots the simulated I-V curve indicating that the breakdown voltage was ~ 31.2 V.

The testing was performed with a single SPAD on chip 1 and a passive quenching circuit (PQC) on chip 2. The two dies were connected by wire bonding as shown in Fig. 5. Figure 6 schematically illustrates the experimental setup. The experiments were performed in a bark environment and the distance between SPAD chip and laser diode is 15 mm. In order to accurately measure the photocurrent, we used an optical chopper together with a lock-in amplifier to measure dark and light current. The laser (Thorlabs, ML925B45F) power was calibrated with an InGaAs photodiode (Thorlabs,

Figure 5. (a) OM photo of the t est dies of two chip 1 (dingle SPADs) and one chip 2 (Passive quenching circuit, PQC). (b) PQC schematic, where V_{OP} provides excess bias voltage (V_{EX}) and V_Q could decide the quenching resistance value.
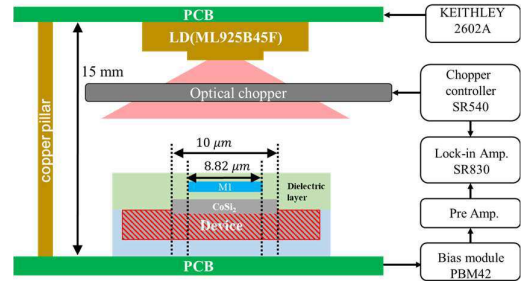
Figure 6. Schematic of the experimental setup in a dark box for the front-side illuminated photocurrent measurements.

SM05PD5A). All optical experiments were performed under front-side illumination.

III. RESULT AND DISCUSSION

Figure 7a shows a semi-log plot of the measured I-V curve and a typical Schottky diode is obtained. The breakdown is also observed and its breakdown voltage, which was defined with the reverse current of 1 μA, was about 31.8 V. The enlarged forward I-V in Fig. 7b was used to extract the ideal factor of ~1.05 and the Schottky barrier height of 0.646 eV which is lower than 1550-nm photon energy of 0.8 eV. The dark current of Schottky SPAD is clearly larger than that of p-n based SPAD, which may cause problems of high dark count rate (DCR) and limited dynamic range.

Figure 4. Simulated reverse-biased current - voltage characteristic in linear and semi-log scale.
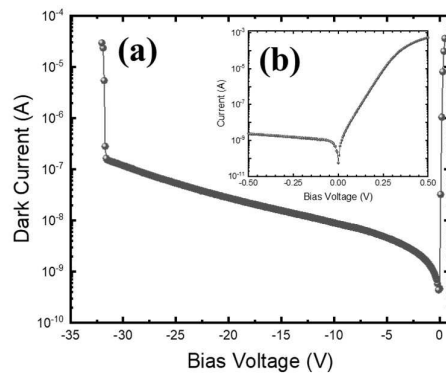
Figure 7. Measured I-V curves in (a) full range, and (b) enlarged around 0 V.

Figure 10. Histogram of inter-arrival time for V_Q = 0.90V, 0.95V, 1.00V, and 1.10V at excess bias = 1.5V.

Figure 8. (a) Net Photocurrent and (b) responsivity versus Reversed bias with ML925B45F 1550nm laser diode operating on Iop = 20, 30, 46 mA.

Figure 8a shows the measured bias-dependent photocurrent under illumination from a 1550-nm continuous-wave diode laser at various intensities. Figure 8b illustrates the responsivity at 15 V of ~60 mA/W, which is pretty small probably due to the inefficient IPE effect for electron injection but showing the potential of Schottky SPAD for SWIR photon counting.

The devices suffered a few issues. For example, as shown in Fig. 9, the measured DCR were not only dependent on operating high voltage (V_{OP}) but also sensitive to quenching resistance which is controlled by quenching NMOSFET gate voltage (V_Q). This also occurred in inter-arrival time (IAT) measurement. Figure 10 shows the IAT histograms. The IAT histogram followed poissonian distribution and the after-pulsing probability was quite low.

In order to focus on the optical responsivity for Schottky-barrier based SPAD, V_{EX} and V_Q was respectively fixed at 1.0 V and 1.048 V and average deadtime was 80 ns in the following measurements. Figure 11a shows the

intensity-dependent photocurrent under illumination from a 1550-nm continuous-wave diode laser measured by the lock-in method mentioned above. The result indicates early photocurrent saturation occurred at the illuminated power of 0.21 pW although the laser diode was operated in linear regime. Figure 11b shows measured intensity-dependent dark and light counts with their respective standard deviations on left y-axis and photon detection probability (PDP) on right y-axis. The Schottky-SPAD has a lower dynamic range because the light count non-linearly observed at the laser operating current of 0.8 mA. In the photocurrent and the PDP measurements, the results evidence the optical response to SWIR photons with early saturation. However, the physical mechanism of the saturation is unknown. We reckon that may be caused by the interface trap related dynamics, such as the recombination and generation at the metal-semiconductor interface. Further works are certainly needed to investigate its Geiger mode operation and quenching behavior to explain
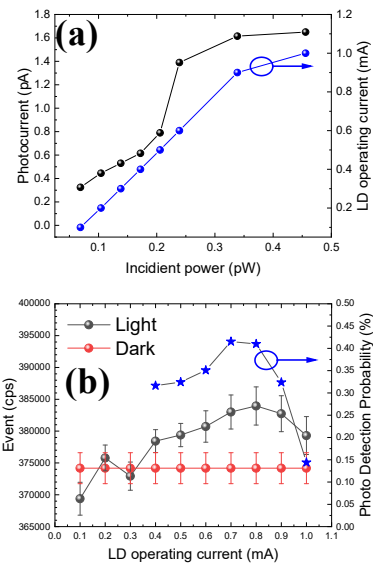
Figure 11. (a)Photocurrent and laser diode (LD) operating current versus incident power and (b) dark and light counting and photon detection probability versus LD operating current

Figure 9. Dark count rate versus excessed bias for V_Q = 0.90V, 0.95V, 1.00V, and 1.10V.

the unusual dependence of DCR/PDP on the excess bias and the quenching resistance.

IV. CONCULSION

We demonstrate a SWIR photon counting detector with Schottky-barrier based SPAD. The devices were fabricated in a conventional CMOS process with integrated passive quenching circuits. We have successfully realized silicon-only SWIR detector. For 1550-nm photons, the Schottky SPAD obtained the responsivity of ~60 mA/W at 15 V and the PDP of ~0.35%. It unfortunately has a very low dynamic region partly because of the intrinsic high dark current of the Schottky diode. Among the bias-dependent characteristics, the quenching resistance and excess bias voltage could play a key role in the junction biasing condition as well as the measured DCR.

REFERENCES

[1] F. Villa, F. Severini, F. Madonini, and F. Zappa, "SPADs and SiPMs Arrays for Long-Range High Speed Light Detection and Ranging (LiDAR)," *Sensors*, vol. 21, pp. 3839, Apr. 2021.

[2] F. Zhao, H. Jiang, and Z. Liu, "Recent development of automotive LiDAR technology, industry and trends," *Proc. SPIE*. Vol. 11179, pp. 1132-1139, Aug. 2019.

[3] D. C. Dumas, J. Kirdoda, R. Millar, P. Vines, and K.Kuzmenko, "High-efficiency Ge-on-Si SPADs for short wave infrared," *Proc SPIE*, vol. 10914, pp. 389-395, Feb. 2019.

[4] F. Signorelli, F. Telesca, E. Conca, A. Della Frera, A. Ruggeri, A. Giudice, and A. Tosi, "InGaAs/InP SPAD detecting single photons at 1550 nm with up to 50% efficiency and low noise," IEDM, pp. 20-3, Dec. 2021.

[5] T. Staffas, M. Brunzell, S. Gyger, L. Schweickert, S. Steinhaurt, and V. Zwiller, "3D scanning quantum LIDAR," CLEO: Applications and Technology, pp. AM2K.1, May 2022.

[6] W. Diels, M. Steyaert, and F. Tavernier, "Schottky diodes in 40nnm bulk CMOS for 1310 nm high-speed optical receivers," OFC, pp. 1-3, Mar. 2017.

[7] T. Maeda, M. Okada, M. Ueno, Y. Yamamoto, and M. Horitia, and J. Suda, Appl. Phys. Express, vol. 9, pp. 091002, Aug. 2016.

A Burst Mode 20 Mfps Low Noise CMOS Image Sensor

Xin Yue, Eric R. Fossum
Thayer School of Engineering, Dartmouth College, Hanover, NH, USA
Contact: xin.yue.th@dartmouth.edu – STUDENT PAPER

Abstract

This paper presents an ultra-high-speed CMOS image sensor utilizing charge-sweep transfer gate technology. This technology eliminates the need for advanced process customization and enables total noise reduction by optimizing the pixel conversion gain.

We have implemented a test chip with a resolution of 64 (columns) by 64 (rows) in a standard 180 nm process and characterized part of its performance. Our testing results demonstrate agreement with theoretical analysis and simulation in areas such as charge transfer time, conversion gain, and readout noise.

Introduction

High-speed CMOS image sensors are widely used in various scientific, industrial, and medical applications. While the current state-of-the-art image sensors reported in literature achieve over 100 million frames per second (Mfps) through process customization [1,2,3,4], this approach can be prohibitively expensive for small-volume customers, and accessing fabrication process modifications can be challenging, especially during the COVID-19 pandemic. Moreover, high-speed CMOS image sensors are prone to higher noise due to the trade-off between the design requirement for fast readout speed, which favors smaller capacitance, and lower thermal noise, which necessitates larger capacitance. In [5], the lowest state-of-the-art input-referred noise was reported to be 8.4 e- rms.

This paper introduces a methodology for optimizing charge transfer time and the concept of charge-sweep transfer gates. We demonstrate that these techniques can be implemented using a standard 180 nm process and enable a CMOS image sensor to achieve over 20 Mfps frame rate. We also discuss optimizations for the floating diffusion, in-pixel correlated double sampling (CDS) circuitry, and memory array, which further reduced the input-referred noise without degrading the frame rate.

The structure of the paper is as follows: in the first section, we describe our approach to designing the photodiode and transfer gates. Then, we discuss the circuitry for in-pixel CDS and the memory array. Finally, we present the results of the characterization and analyze their limitations.

Photodiode Optimization

From the perspective of charge transportation, it is well-known that electrons can achieve a higher velocity in a strong electrical field. To leverage this, we propose creating a lateral electrical field along the charge transfer direction in the pixel. Equation 1 from [7,9] provides a simplified relationship between the maximum electrostatic potential (ψ) in a photodiode, the elementary charge (q), the doping concentration of the photodiode (N_D), the doping concentration of the substrate (N_A), and the photodiode half width (X_n). By adjusting X_n parabolically, a constant electrical field can be established [1,7] from the tip of the photodiode to the transfer gate, as described by Equation 2, where x and y stand for the coordinates of the photodiode finger. To achieve an optimal trade-off between pixel fill factor and charge transfer time, we implement and simulate several different photodiode designs in TCAD, as depicted in Figure 1 [8]. Our results, summarized in Table 1, show that the E800 (800 V/cm) design outperforms the others. Therefore, we select this design for the rest of the pixel finger design.

$$\psi_{max} \approx \frac{q \cdot N_D \cdot X_n^2}{2 \cdot \varepsilon_0 \cdot \varepsilon_r}\left(1 + \frac{N_D}{N_A}\right) \quad (1)$$

$$y = -\frac{q \cdot N_D \cdot x^2}{2 \cdot E \cdot \varepsilon_0 \cdot \varepsilon_r}\left(1 + \frac{N_D}{N_A}\right) + C_0 \quad (2)$$
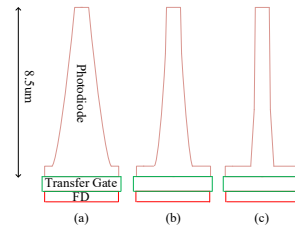

Figure 1. Sample pixel layouts

Charge Transfer Time of Different Photodiodes					
CTE	90%	99%	99.5%	99.9%	Unit
E400	10.2	41.2	51.3	75.2	ns
E500	0.67	17.4	25.8	47.0	ns
E600	0.67	5.01	11.0	28.3	ns
E700	0.74	1.51	5.16	19.3	ns
E800	0.82	1.18	3.78	15.5	ns

Table 1. Charge transfer time of different photodiode designs

Charge-sweep Transfer Gate

One may observe that the transfer gate in the sample pixels depicted in Figure 1 has a comparable width to the pixels. As a result, when dealing with large pixels, such as 20 μm * 20 μm, the transfer gate width is also around 20 μm. This will result in a considerable floating diffusion node area [8] and a reduction in pixel conversion gain. As shown in [6], the conversion gain is estimated to be less than 10 μV/e- for a 20 μm pixel. To address this issue, we propose utilizing charge-sweep transfer gates (TX3, TX2, and TX1), as depicted in Figure 2. Each gate features a smaller geometry size than the prior one, resulting in a smaller floating diffusion node, as highlighted in the red rectangle.
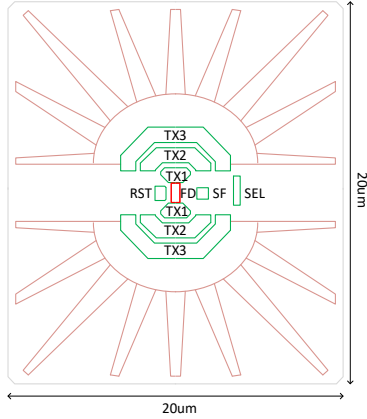
Figure 2. High-speed pixel layout based on charge-sweep transfer gate

In a 180 nm process, the typical gap between two poly gates is 0.2 μm ~ 0.3 μm. We developed two timing sequences to achieve complete charge transfer from the photodiode to the floating diffusion node without using a double poly gates process or implementing special doping beneath the transfer gates. Figure 3 depicts the two timing sequences.
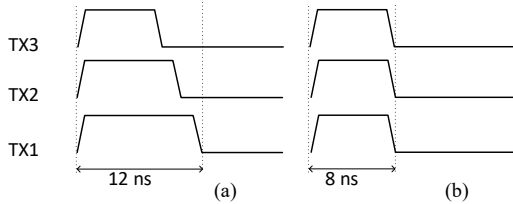

Figure 3. TX gates timing for charge-sweep transfer gates

In timing sequence a), TX1's On voltage is slightly higher than that of TX2, and TX2's On voltage is slightly higher than that of TX3. At the start of the charge transfer, all three gates, TX1, TX2, and TX3, are turned on. As the charge transfer comes to an end, TX3 is the first to turn off, followed by TX2, and finally, TX1. Considering the rise and fall time of the TX pulses, the complete charge transfer sequence takes 12 ns in simulation.

In timing sequence b), the On voltage of TX1 is considerably higher than that of TX2, and the On voltage of TX2 is significantly higher than TX3. This removes the potential barrier between the adjacent gates while they are turned on. Initially, all three gates, TX1, TX2, and TX3, are switched on during charge transfer, and then all three gates are turned off simultaneously when charge transfer is complete. The complete charge transfer sequence is simulated to take only 8 ns.

Floating Diffusion Node Optimization

For a typical floating diffusion, self-alignment technology allows the N+ implant to fully cover the transfer gate and floating diffusion, leaving no gap in between, which facilitates charge transfer. However, for pixels with charge-sweep transfer gates, complete charge transfer can only be achieved after all gates have been fully turned off. Therefore, it is safe to move the floating diffusion away from the TX1 gate and create a gap in between [10,11], as depicted in Figure 4. This can effectively reduce the parasitic capacitance overlap between the floating diffusion node and TX gate and further improve the conversion gain.
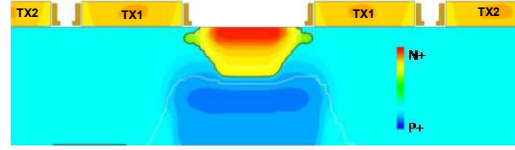
Figure 4. Cross-section of the doping profile of floating diffusion node

Pixel Source Follower

The analysis of the capacitance distribution at the floating diffusion node was conducted, as depicted in Figure 5, which revealed that the gate-to-drain capacitance ($C_{fd\_sf\_gd}$) and gate-to-ground capacitance (C_{fd_gnd}) of the source follower dominated. To enhance the pixel conversion gain, the high-conversion-gain (HCG) variant removed the lightly-doped drain (LDD) on the drain side [12] and decreased the gate length from 0.6 μm to 0.3 μm, as shown in Figure 6. TCAD simulations demonstrate that the modification resulted in an increase in the pixel conversion gain from 138 μV/*e*- to 174 μV/*e*-.
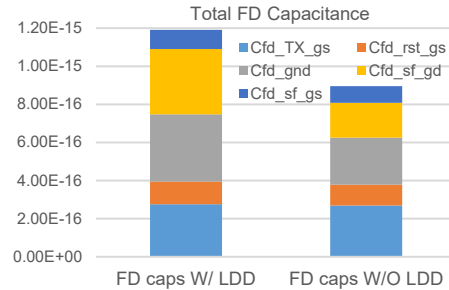
Figure 5. Capacitance distribution of baseline pixel (left) and HCG pixel (right)
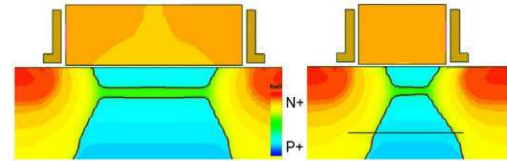
Figure 6. Cross-section of the doping profile of SF in baseline pixel (left) and HCG pixel (right)

In-Pixel CDS Circuitry

Similar to many CMOS image sensors, the flicker and thermal noise of the pixel's first stage source follower (SF) typically dominate the input-referred noise. Without altering the standard fabrication process or incorporating advanced interface passivation, correlated double sampling (CDS) remains a useful method for decreasing low-frequency thermal noise and flicker noise.

To account for the voltage gain attenuation introduced by the CDS circuit, we are implementing the circuit shown in Figure 7 in this pixel [13]. Specifically, we are placing the C_{SH} at the output of the first-stage source-follower instead of the input of the second-stage source-follower, as described in [1,5]. This configuration allows us to reduce the voltage attenuation in the signal chain to $C_{CDS}/(C_{CDS}+C_P)$, where C_{CDS} stands for the AC CDS capacitor, and C_P is the parasitic capacitor.

The following section will provide details of the 1.8 V thin gate sample/hold capacitor bank. To protect the 1.8 V thin gate devices in a 3.3 V environment, the V_{RST} voltage is

isolated from the V_{DDpix} and can be adjusted autonomously, with V_{RST} usually set to $1.8+V_{GS\_SF2}$. This configuration guarantees that the SF2's maximum output voltage stays below 1.8 V.
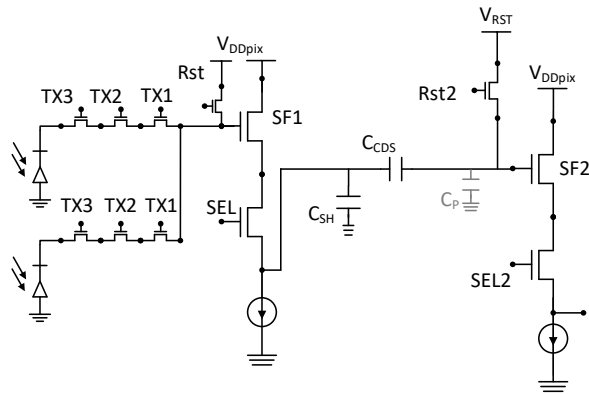

Figure 7. In-pixel CDS circuit

Sample/Hold Capacitor Unit

For design simplicity and durability in a 3.3 V operating environment, it is preferable to use thick gate 3.3 V devices. However, the difference in dielectric layer thickness leads to a lower capacitance density of 3.3V NMOS capacitors, which typically ranges from 0.25 to 0.5 of that of 1.8 V thin gate NMOS capacitors, and an increase in thermal noise. To overcome this challenge, this pixel utilizes 1.8 V NMOS capacitors in the sample and hold capacitor bank.

To achieve a higher capacitance density, a custom Metal-1 (M1) Metal-Oxide-Metal (MOM) capacitor is installed on top of the poly gate of the NMOS capacitor. Moreover, a Metal-2 (M2) layer acts as a shielding layer positioned above the M1 MOM capacitor, as depicted in Figure 8. By implementing this design, we were able to fit 108 units of sample and hold capacitors (each with a capacitance of 78 fF) into a 52 μm pixel in the final layout.
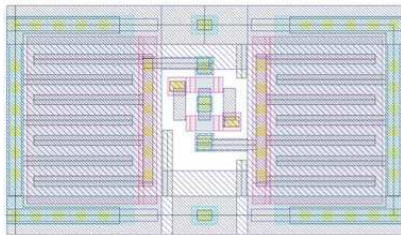
Figure 8. The layout of in-pixel Sample/Hold unit

Top Chip Power Distribution

One of the challenges involved in designing a burst mode CMOS image sensor pertains to the power distribution network. In particular, during the pixel resetting phase, a significant amount of instantaneous current is necessary to reset both the floating diffusion node and CDS capacitors. If the supply network has high resistance, temporary collapses on the supply rails may occur and take time to recover. To reduce routing resistance, the power and reference rails associated with pixels are placed on the top thick metal layer in the layout and are star-connected to all four sides of the pad ring. Figure 9 highlights these connections, which are enclosed by red boxes.
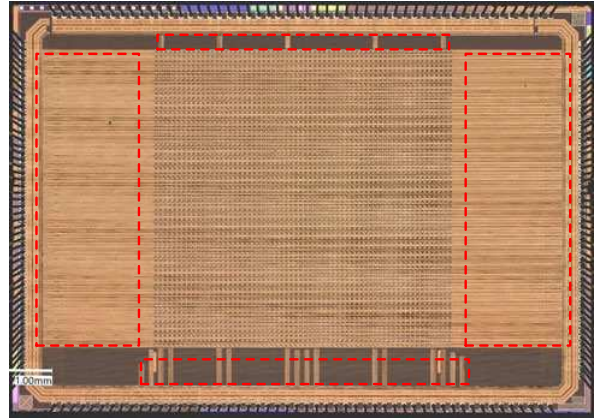
Figure 9. Microscope Image of the sensor

Test System and Measurement

Despite demonstrating in [6] through TCAD simulation that the sensor is capable of operating at 20 Mfps, the current prototype test system is constrained by the hardware capabilities of the FPGA, prototype PCB, and chip carrier, which restricts reliable operation to a maximum of 15.6 Mfps. The prototype system utilizing a CPGA-208 package and a zero-insertion-force (ZIF) socket introduces parasitic inductances that cause significant ringing on the power supply during pixel reset operations. This ringing can result in CDS errors and increase noise if the power supply and reference voltage have not fully settled before the end of CDS sampling. Increasing the CDS reset pulse (Rst2 in Figure 7) width can suppress this artifact, but it also reduces the frame rate of the sensor. Hence, to achieve optimal noise performance, we conducted the remaining measurements at a frame rate of 4 Mfps.

The total output noise was measured for both the baseline pixels and high-conversion-gain (HCG) pixels, as depicted in Figure 10. The baseline pixels exhibited a noise level of 10.6 DN at the sensor output, which is equivalent to 12 $e-$ rms at the input. In contrast, the HCG pixels were expected to have higher flicker noise due to the smaller in-pixel source follower gate area. However, the short CDS period (80 ns) canceled out the majority of the noise, resulting in a total output noise of 11.6 DN, which is equivalent to 7.6 $e-$ rms at the input, as shown by the silicon measurement.
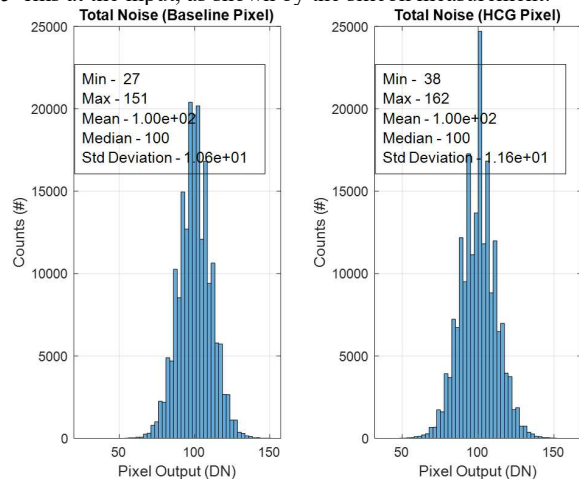

Figure 10. Total noise measurement result for baseline pixel (left) and HCG pixel (right)

In Figure 11, the Photon-Transfer-Curve (PTC) was measured for both pixel types. The measured data, adjusted by the voltage gain of 0.485 V/V across the entire signal chain and ADC LSB 38 μV/DN, indicates that the baseline pixel has a conversion gain of 69 μV/*e-*, whereas the HCG pixel has a conversion gain of 119 μV/*e-*.
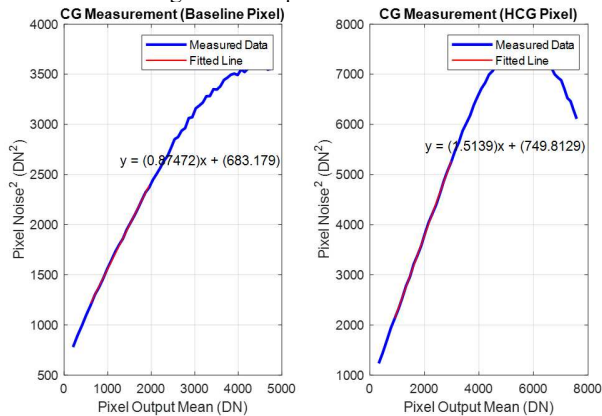

Figure 11. *Conversion-Gain measurement result for baseline pixel (left) and HCG pixel (right)*

The image lag test was conducted on both the baseline and HCG pixels, and Figure 12 shows the results. The measurements reveal that the baseline pixel has a negligible lag (<0.1%). On the other hand, the HCG pixel displays an approximately 3% lag, which is due to overflow at the floating diffusion node.
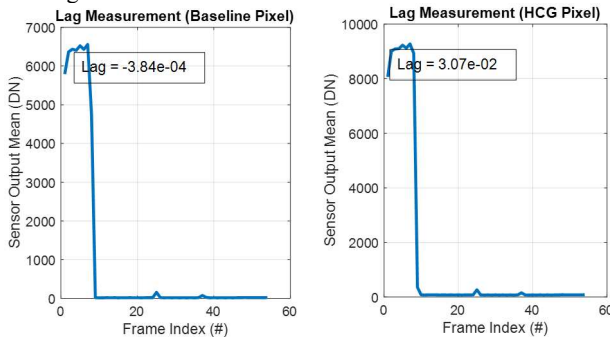
Figure 12. *Lag measurement result for baseline pixel (left) and HCG pixel (right)*

Conclusion

The initial characterization results indicate that the use of a charge-sweep transfer gate can enhance the pixel conversion gain and decrease the input-referred noise. Unfortunately, due to time constraints, certain measurements, such the quantum efficiency, were left incomplete before the paper submission deadline. Nonetheless, we aim to present supplementary test findings in future research.

Acknowledgments

References

[1]   Suzuki, M., Sugama, Y., Kuroda, R., & Sugawa, S. (2020). Over 100 million frames per second 368 frames global shutter burst CMOS image sensor with pixel-wise trench capacitor memory array. *Sensors*, *20*(4), 1086.

[2]   Tochigi, Y., Hanzawa, K., Kato, Y., Kuroda, R., Mutoh, H., Hirose, R., Tominaga, H., Takubo, K., Kondo, Y., & Sugawa, S. (2013). A global-shutter CMOS image sensor with readout speed of 1-tpixel/s burst and 780-mpixel/s continuous. The IEEE Journal of Solid-State Circuits, 48(1), 329–338.

[3]   Dao, V. T. S., Ngo, N., Nguyen, A. Q., Morimoto, K., Shimonomura, K., Goetschalckx, P., ... & Etoh, T. G. (2018). An image signal accumulation multi-collection-gate image sensor operating at 25 Mfps with 32× 32 pixels and 1220 in-pixel frame memory. Sensors, 18(9), 3112.

[4]   Mochizuki, F., Kagawa, K., Okihara, S. I., Seo, M. W., Zhang, B., Takasawa, T., ... & Kawahito, S. (2016). Single-event transient imaging with an ultra-high-speed temporally compressive multi-aperture CMOS image sensor. Optics express, 24(4), 4155-4176.

[5]   Wu, L., San Segundo Bello, D., Coppejans, P., Craninckx, J., Süss, A., Rosmeulen, M., Wambacq, P., & Borremans, J. (2018). Analysis and design of a CMOS ultra-high-speed burst mode imager with in-situ storage topology featuring in-pixel CDS amplification. *Sensors*, *18*(11), 3683.

[6]   Yue, X., & Fossum, E. R. (2023). Simulation and design of a burst mode 20Mfps global shutter high conversion gain CMOS image sensor in a standard 180nm CMOS image sensor process using sequential transfer gates. Electronic Imaging, 35, 1-5.

[7]   Takeshita, H., Sawada, T., Iida, T., Yasutomi, K., & Kawahito, S. (2010, January). High-speed charge transfer pinned-photodiode for a CMOS time-of-flight range image sensor. In Sensors, Cameras, and Systems for Industrial/Scientific Applications XI (Vol. 7536, pp. 235-243). SPIE.

[8]   Cao, X., Gäbler, D., Lee, C., Ling, T. P., Jarau, D. A., Tien, D. K. C., ... & Bold, B. (2015). Design and optimisation of large 4T pixel. In Proc. Int. Image Sensor Workshop (IISW) (pp. 112-115).

[9]   Park, S., & Uh, H. (2009). The effect of size on photodiode pinch-off voltage for small pixel CMOS image sensors. Microelectronics Journal, 40(1), 137-140.

[10] Chen, S., Ma, J., Hondongwa, D. B., & Fossum, E. R. (2017). High conversion-gain pinned-photodiode pump-gate pixels in 180-nm CMOS process. IEEE Journal of the Electron Devices Society, 5(6), 509-517.

[11] Ma, J., & Fossum, E. R. (2015). A pump-gate jot device with high conversion gain for a quanta image sensor. IEEE Journal of the Electron Devices Society, 3(2), 73-77.

[12] Kusuhara, F., Wakashima, S., Nasuno, S., Kuroda, R., & Sugawa, S. (2016). Analysis and reduction technologies of floating diffusion capacitance in CMOS image sensor for photon-countable sensitivity. ITE Transactions on Media Technology and Applications, 4(2), 91-98.

[13] De Wit, Y., Walschap, T., & Cremers, B. (2010). U.S. Patent Application No. 12/766,798.

Towards Infrared Spectral Extension of CMOS Image Sensors

Kaitlin M. Anagnost, Xiaoxin Wang, Jifeng Liu, and Eric R. Fossum
Thayer School of Engineering, Dartmouth College, Hanover, NH, USA
Contact: Kaitlin.Anagnost.TH@dartmouth.edu

Abstract

A new structural integration scheme for p-type infrared (IR) absorbers directly on silicon (Si)-based CMOS image sensors with type-II band alignment is proposed and explored by calculation as an alternative to HgCdTe and other hybridized IR detectors. While HgCdTe is a material of choice for many IR detectors, its challenging manufacturing process and thermal reliability, among other factors, prove detrimental to some applications. Si-based sensors with directly deposited IR absorbing layers may be a suitable alternative. The band structures for the IR absorbing materials $Ge_{0.89}Sn_{0.11}$ and $In_{0.1}Ga_{0.9}Sb$ on Si are calculated and analyzed. Detector parameters including dark current, target wavelength, quantum efficiency (QE), and others, are also calculated to explore this approach in concert with separate experimental fabrication and measurement.

Keywords—CMOS image sensor, Non-visible, Spectral extension, IR.

Introduction

There is a desire to transition towards Si-based IR sensors due to easier readout, potentially lower readout noise, and other circuit integration on-chip. Further, they are expected to avoid the manufacturing, scalability, and thermal reliability challenges HgCdTe and other hybrid sensors face. However, without an IR absorption layer, Si has a weak IR response because of its 1.12 eV bandgap [1]. IR imaging has many uses, including in light detection and ranging, security, medical, and more.

A new structural integration scheme for p-type infrared (IR) absorbers directly on silicon (Si)-based CMOS image sensors (CIS) with type-II band alignment is proposed and explored by calculation as an alternative to state-of-the-art hybrid IR detectors. In the proposed sensor, photons strike a p-type IR absorption layer that forms a type-II heterostructure on Si, generating photoelectrons that diffuse into the pixel's n-type Si storage well, leaving the holes in the p$^+$ Si pinning region. Depositing the IR absorber directly on silicon negates the need for metal hybrid bonds or bump bonds typically used in many IR detectors. Aside from reliability concerns during thermal cycling and yield, a metal interconnect implies a 3-T type readout with residual kTC noise and potential lag. Direct injection of photoelectrons into Si permits low-noise 4-T readout.

$Ge_{0.89}Sn_{0.11}$ and $In_{0.1}Ga_{0.9}Sb$ are considered as possible materials for the IR absorption layer. Like HgCdTe, GeSn has a direct, tunable bandgap, but is CMOS- and Si-compatible, and thus more scalable [2]–[4]. Experimental results in [5] verify proof-of-concept with 100 mA/W responsivity at ~2 μm compatible with back-end-of-line CMOS processing. III-V materials like InGaSb also have direct bandgaps [6] and are easier to manufacture than II-VI compounds [7].

To advance this approach towards use on a CIS, the electron diffusion and subsequent readout is first optimized via an interdigitated structure. Deposition layer thickness, minimum interdigitated structure width, target wavelength, quantum efficiency, and dark current are then calculated to further characterize the proposed device.

Band Structure Calculation and Analysis

Some photoelectrons will diffuse to locations that won't be readout if the IR absorber is deposited over the entire p+ Si pinning layer commonly used in frontside-illuminated devices. Therefore, an interdigitated Si layer with p+ and n type regions, equivalently a partial pinning layer, is considered as shown in Figure 1. This can be implemented with an interdigitated mask for p+ pinning layer implantation.
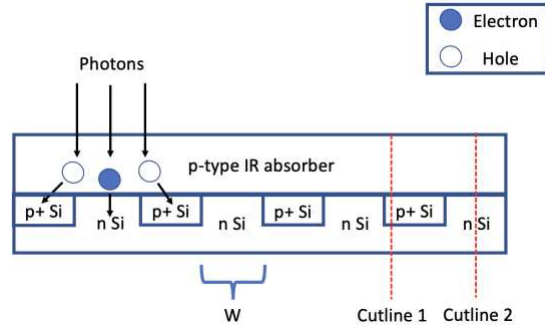

Figure 1. IR-absorbing layer on interdigitated silicon layer.

The IR-absorber/Si band structures are calculated using the Anderson approach, which utilizes the vacuum level to find the valence band offset. The difference in conduction band energies ΔE_c is the difference between the Si and IR absorbers' affinities, $q\chi_{Si}$ and $q\chi_{absorber}$, respectively. Using the direct bandgap of the IR absorber and indirect bandgap of Si (1.12 eV), we find the difference in IR absorber and Si bandgaps ΔE_g and thus valence band energies ΔE_v:

$$\Delta E_v = \Delta E_g - \Delta E_c \qquad (1)$$

Assuming Si's vacuum level is $q\chi_{Si} = 4.05$ eV, since $E_c = E_v + E_g$, we can determine Si's E_c and E_v, and use $\Delta E_v, \Delta E_c,$ and ΔE_g to find those of the absorbers.

To calculate the Fermi level E_F, the density of states in the conduction band N_c and valence band N_v are first found using (2.13a) and (2.13b) on pp. 51 in [8]. The intrinsic Fermi level E_i and intrinsic carrier concentration n_i are then determined via (2.36) on pp. 62 and (2.21) on pp. 55 in [8], respectively. Finally, E_F is calculated using 2.38a and 2.38b on pp. 63 in [8].

The depletion region widths on each side of the junction are found next using (143a) and (143b) in [9]. The Debye lengths L_D are subsequently determined and compared to the depletion widths to validate the analyses.

The band bending present in the IR absorbers and Si are found next. The sum of the interface potentials for the absorber and Si, $q\psi_{absorber} + q\psi_{Si} = q\psi_{bi}$ and with 0 V applied, are related by (15a) and (15b) on pp. 82 of [9]. The Fermi levels of the materials are then aligned at 0 eV for simplicity.

The equilibrium band diagrams for the crosscuts of Fig. 1 were calculated with the results using $Ge_{0.89}Sn_{0.11}$ and $In_{0.1}Ga_{0.9}Sb$ shown in Fig. 2 and 3, respectively. As illustrated, the p+ Si regions in Figure 2a and 3a serve as a potential barrier against photoelectron transport but collect holes generated in the IR absorber. Conversely, the p-type IR absorber/n-type heterostructure allows electron diffusion to the n-well, as depicted in Fig. 2b and 3b. The $Ge_{0.89}Sn_{0.11}$/n Si conduction band structure has no barrier for electron diffusion from $Ge_{0.89}Sn_{0.11}$ to n Si, while the $In_{0.1}Ga_{0.9}Sb$/Si conduction band has a small barrier that allows electrons to tunnel to the n Si well. $Ge_{0.89}Sn_{0.11}$ is thus more suited to this design as an IR absorbing layer.

(b)
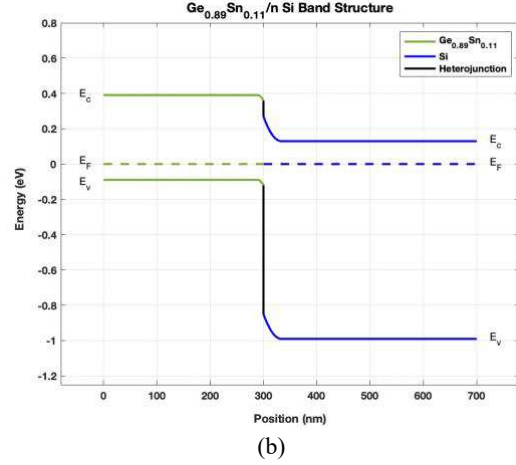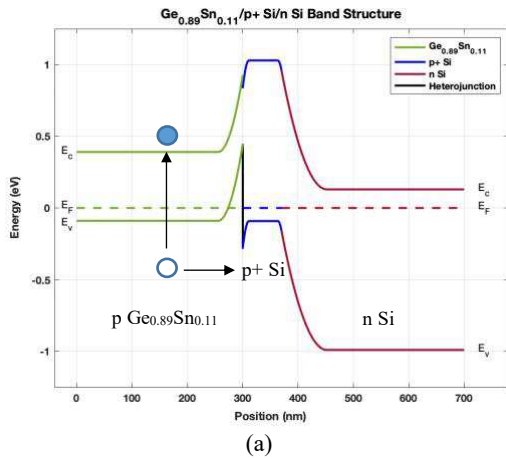
Figure 2. Calculated band structure for $Ge_{0.89}Sn_{0.11}$ for (a) Cutline 1 and (b) Cutline 2 in Figure 1.

(a)

(b)
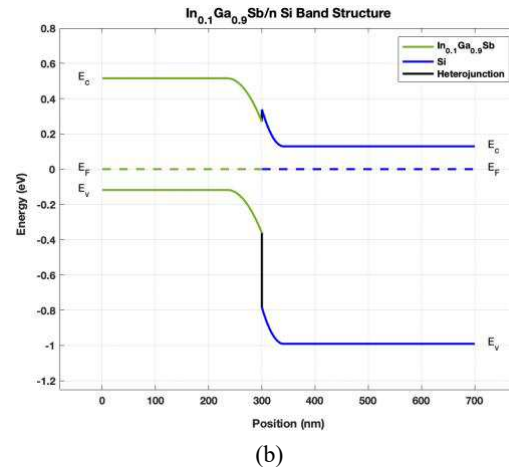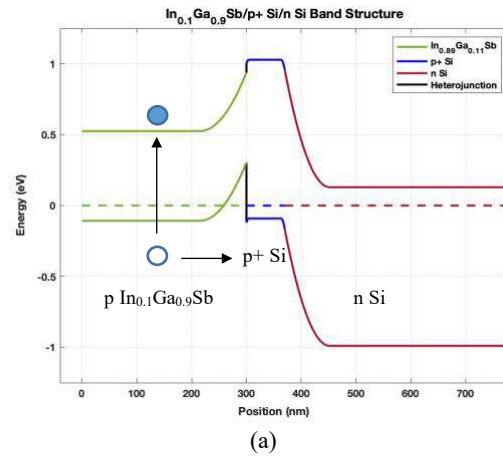
Figure 3. Calculated band structure for $In_{0.1}Ga_{0.9}Sb$ (a) Cutline 1 and (b) Cutline 2 in Figure 1.

(a)

QE Estimation

Deposition layer thickness, minimum interdigitated structure width, target wavelength, and QE are evaluated next to obtain a greater understanding of the device.

The ideal deposition thickness of the materials are first found using Beer-Lambert's law [10]

$$I = I_0 e^{-\alpha z} \tag{2}$$

where z is the material's depth, α is the absorption coefficient, and I_0 is the initial intensity.

The ratio of the amount of light absorbed and readout to the total amount of light incident on the material, or external quantum efficiency (EQE), is

$$EQE = A \times CP \tag{3}$$

where A is the amount of incident light absorbed in the device and CP is the collection probability, or probability that the light will be readout and is given by 8.13 in [11] as

$$CP = e^{-x/L_D} \tag{4}$$

where x is the distance from the depletion region and L_D is the minority carrier diffusion length. Light absorbed within the depletion region has a 100% probability of being collected in the absence of a potential barrier. However, the CP for light outside the depletion region exponentially decreases with distance from the depletion region. As a result,

$$EQE = A_{out} \times CP + A_{depl} \tag{5}$$

where A_{out} and A_{depl} are the percentages of light absorbed outside and inside of the depletion region, respectively. A_{depl} is given by

$$A_{depl} = e^{-\alpha(z - w_{depl})} - e^{-\alpha z} \tag{6}$$

with no potential barrier present. In the presence of a potential barrier, A_{depl} becomes $A_{depl} \times T$, the tunneling probability.

Since the CP varies as a function of length, A_{out} must be calculated for each position used. Using (2),

$$A_{out} = \sum_{i=1}^{i=1000} (e^{-\alpha z_{i-1}} - e^{-\alpha z_i}) \tag{7}$$

where i is the bin number, $z_i = (z - w_{depl})(i)(\Delta z_i)$, $\Delta z_i = 0.001$, and w_{depl} is the depletion region width. The absorption in each bin is subtracted from the previous one so the value isn't cumulative. The EQE is then found, with the results displayed in Table 1.

The ratio of the number of carriers readout to the number of photons absorbed, or the internal quantum efficiency (IQE), can also be compared between the materials. For $Ge_{0.89}Sn_{0.11}$, the IQE is simply given by the collection probability CP, while the IQE for $In_{0.1}Ga_{0.9}Sb$ is CP × T due to the potential barrier.

The refractive index is subsequently calculated, and the minimum interdigitated structure width is found using Snell's law, with the results shown in Table 1.

Dark Current Estimation

The dark current is critical, especially in low-light conditions in which the photo signal may become overwhelmed by the noise. It is first calculated at 300 K and using 100 mV reverse bias. Only the variables and equations for electrons will be stated to avoid redundancy, with the equivalent expression for holes left to the reader.

First the electron concentration n of the IR absorber is determined from 2.16a in [20]:

$$n = N_C e^{(E_F - E_c)/kT} \tag{8}$$

The change in electron concentration with respect to time, $\frac{\partial n}{\partial t}$, at the IR absorber/Si interface is given by (5) in [12] and is the dark current in this device

$$\frac{\partial n}{\partial t} = \frac{1}{q}\frac{d}{dx}J_n - R_{bb} - R_{nt} \tag{9}$$

where J_n is the electron current density, R_{bb} is the band-to-band recombination term, and R_{nt} is the transition rate from traps to the conduction band. From (8) in [12]:

$$J_n = qn\mu_n E + qD_n \nabla n \tag{10}$$

where μ_n is the electron mobility, E is the electric field, and D_n is the diffusion constant in an n-type semiconductor.

The band-to-band recombination coefficient is

$$R_{bb} = B_r(np - n_i^2) \tag{11}$$

where B_r is the recombination coefficient [13], [14] (19), (Appendix C). The transition rate from traps to the conduction band, R_{nt}, from [13] (8, 10), is

$$R_{nt}(x) = \int_{E_v}^{E_c} r_{nt}(E_t, x)dE_t \tag{12}$$

E_t is the trap energy and r_{nt} is the transition rate from traps at particular energy levels to the conduction band [13] (9, 11)

$$r_{nt}(E_t) = c_n n(1 - f_t)D_t - e_n f_t D_t \tag{13}$$

where c_n is the electron capture coefficient, f_t is the occupation function, D_t is the defect density of states, and e_n is the electron thermal emission rate. $\frac{\partial n}{\partial t}$ is then found, with the results shown in Table 1.

Results

As illustrated in Table 1, $Ge_{0.89}Sn_{0.11}$'s target wavelength extends farther into the IR regime due to its smaller bandgap. The minimum interdigitated width for the structures isn't a concern because of the IR absorbing layer's thickness. $In_{0.1}Ga_{0.9}Sb$'s EQE and IQE are low because the potential barrier in its conduction band

hinders electron transport to the Si. $Ge_{0.89}Sn_{0.11}$ has more dark current because its smaller bandgap allows carriers to jump into the conduction and valence bands with less energy than in materials with larger bandgaps like $In_{0.1}Ga_{0.9}Sb$. To achieve 1 pA/cm² dark current with 100 mV reverse bias, $Ge_{0.89}Sn_{0.11}$ and $In_{0.1}Ga_{0.9}Sb$ would need to be cooled to 155 K and 185 K, respectively, assuming mid-gap defect state excitation at the heterojunction interface dominates the thermal generation process over that of band-to-band processes.

Parameter	$Ge_{0.89}Sn_{0.11}$	$In_{0.1}Ga_{0.9}Sb$
IR Absorber Bandgap	0.48 eV	0.63 eV
Wavelength Range	0.4-2.1 um	0.4-2.0 um
IR Absorber Thickness	100 nm	100 nm
External Quantum Efficiency	23%	0.9%
Internal Quantum Efficiency	~100%	~10%
Minimum Interdigitated Width	48 nm	52 nm
Eff. Trap Density incl. heterojunction interface	$1x10^{17}$/cm³	$1x10^{16}$/cm³
Dark current at 300 K, 100 mV reverse bias	12.6 mA/cm²	0.22 mA/cm²
Temp. for 1 pA/cm^2 dark current	155 K	185 K
Potential Readout Noise (4T config.)	<5e- rms	<5e- rms

Table 1. *Simulated parameters of the proposed detector.*

Discussion

From these calculations, a better understanding of the detector is formed. The results indicate that $Ge_{0.89}Sn_{0.11}$ is more promising than $In_{0.1}Ga_{0.9}Sb$ to allow photoelectron diffusion into the Si. $Ge_{0.89}Sn_{0.11}$ extends the detector's spectral responsivity farther into the IR range with higher QE, but at the cost of higher dark current. Future work includes depositing the $Ge_{0.89}Sn_{0.11}$ on Si-based CMOS image sensors and testing them. If successful, the widespread use of Si-based IR detectors will be closer to being in reach.

Acknowledgments

References

[1] A. Rogalski, "HgCdTe infrared detector material: History, status and outlook," *Reports on Progress in Physics*, 2005, doi: 10.1088/0034-4885/68/10/R01.

[2] D. Zhang *et al.*, "High-responsivity GeSn short-wave infrared p-i-n photodetectors," *Applied Physics Letters*, 2013, doi: 10.1063/1.4801957.

[3] H. Tran *et al.*, "Si-Based GeSn Photodetectors toward Mid-Infrared Imaging Applications," *ACS Photonics*, 2019, doi: 10.1021/acsphotonics.9b00845.

[4] R. Soref, D. Buca, and S.-Q. Yu, "Group IV Photonics: Driving Integrated Optoelectronics," *Optics and Photonics News*, 2016, doi: 10.1364/opn.27.1.000032.

[5] X. Wang *et al.*, "GeSn on Insulators (GeSnOI) Toward Mid-infrared Integrated Photonics," *Frontiers in Physics*, vol. 7, 2019, Accessed: Dec. 01, 2022. [Online]. Available: https://www.frontiersin.org/articles/10.3389/fphy.2019.00134

[6] S. P. Svensson, W. A. Beck, W. L. Sarney, D. Donetsky, S. Suchalkin, and G. Belenky, "Temperature dependent Hall effect in InAsSb with a 0.11 eV 77 K-bandgap," *Appl. Phys. Lett.*, vol. 114, no. 12, p. 122102, Mar. 2019, doi: 10.1063/1.5081120.

[7] E. H. Steenbergen, C. P. Morath, D. Maestas, G. D. Jenkins, and J. V. Logan, "Comparing II-VI and III-V infrared detectors for space applications," in *Infrared Technology and Applications XLV*, SPIE, May 2019, pp. 299–307. doi: 10.1117/12.2519250.

[8] R. Pierret, *Semiconductor Device Fundamentals*. Addison-Wesley Publishing Company, 1996.

[9] S. M. Sze and K. K. Ng, "Physics of Semiconductor Devices, 3rd Edition - Simon M. Sze, Kwok K. Ng," *Physics of Semiconductor Devices, 3rd Edition.; John Wiley & Sons, Inc.; NJ*, 2007.

[10] "Brown and Arnold - 2010 - Fundamentals of Laser-Material Interaction and App.pdf." Accessed: Feb. 22, 2022. [Online]. Available: https://spikelab.mycpanel.princeton.edu/papers/book02.pdf

[11] "Solar cells—Operating principles, technology and system applications," in *Solar Energy*, 1982, p. 447. doi: 10.1016/0038-092X(82)90265-1.

[12] B. Baert, M. Schmeits, and N. D. Nguyen, "Study of the energy distribution of the interface trap density in a GeSn MOS structure by numerical simulation of the electrical characteristics," *Applied Surface Science*, vol. 291, pp. 25–30, Feb. 2014, doi: 10.1016/j.apsusc.2013.09.022.

[13] M. Sakhaf and M. Schmeits, "Capacitance and conductance of semiconductor heterojunctions with continuous energy distribution of interface states," *Journal of Applied Physics*, vol. 80, no. 12, pp. 6839–6848, Dec. 1996, doi: 10.1063/1.363750.

[14] W. Dou, "High-Sn-content GeSn Alloy towards Room-temperature Mid Infrared Laser," University of Arkansas, Fayetteville.

Chip-level Performance Analysis using Test Element Group Devices for indirect Time-of-Flight CMOS Image Sensor

Seunghyun Lee[1], Jungwook Lim[1], Minhee Son[1], Sungyoung Seo[1], Sunghyuck Cho[1],
Taeun Hwang[1], Yonghun Kwon[1], Youngchan Kim[1], Young-Gu Jin[1], Seok-Ha Lee[2],
Seung-Hun Shin[2], Seunghyun Song[2], Youngsun Oh[1], and JungChak Ahn[1]

[1] System LSI Division, Samsung Electronics Co., Ltd., [2] Foundry Division, Samsung Electronics Co., Ltd.
1, Samseong-ro, Giheung-gu, Yongin-si, Gyeonggi-do, 17113, Korea
Phone: +82-10-9756-4690 E-mail: shyun08.lee@samsung.com

We developed Test Element Group (TEG) module to analyze image characteristics of Indirect Time-of-Flight (iToF) pixel. It is difficult to figure out the cause of image degradation, since iToF pixel has more complicated structure compared to conventional 4T CIS pixel. Therefore, it is needed to analyze iToF sensor not only with the chip-level image, but also with single pixel device. In this paper, we developed TEG device, which consists of electric potential curve, C-V curve, and charge pumping method. In addition, we also showed correlations between TEG and chip-level image results.

Indirect Time-of-Flight (iToF) uses 4-sampling method with Global Shutter (GS) operation to calculate depth information of an image. For this operation, multiple devices including MOSCAP storage are connected in a chain structure. Due to the complex structure, it is difficult to define the cause of image degradation, so device-level analysis is needed. In addition, since both operation and structure of iToF are different from 4T CIS, the approach using conventional TEG cannot be used in our sensor. In this work, we designed TEG devices to figure out image characteristics suitable for iToF sensor.

We propose new TEG device designed for iToF pixel structure and operation. Figure 1 shows unit pixel schematics and timing diagram of iToF. First, turn on the Overflow Gate (OG) during the global reset operation time to empty the photodiode (PD). Then, during the integration time, the photogate (PG) is toggled for modulation, and the photoelectron is transferred by Transfer Gate (TGA), and stored under the Storage Gate (SG). After that, read out the floating diffusion (FD) signal by turning SG OFF and TG1 ON [1].

Figure 2 (a) shows the potential of each device during modulation. Shut-off (S/O) and maximum voltage (Vmax) under each device determines the performance of iToF pixel, for example, the Full Well Capacity (FWC). To obtain these electrical potentials, we applied current bias to the source node, and observed change of the source voltage by increasing TGA or TG1's voltage [2-3], an example is shown in figure 2 (b). The measured S/O voltage and Vmax with different SG voltage is shown in figure 3. We also compared the chip-level FWC result and TEG potential measurement. Both measured chip-level FWC and calculated TEG FWC are increased as SG voltage increases. This states that we can predict FWC with TEG measurement.

Figure 4 shows the capacitance of MOSCAP obtained with conventional C-V curve measurement. The operating frequency of PG is 100 MHz, and SG is held at high voltage to store photoelectron during integration time until readout operation. Therefore, we categorized the source of dark current into PG and SG components. At low frequencies and low voltage, the capacitance decreases with increasing temperature. This indicates that the charge at the interface traps becomes more likely to de-trap with increasing temperature. On the other hand, at high frequencies and high negative voltage, the capacitance increases as the temperature increases because the probability of charge tunneling into the border trap inside the oxide increases. In this way, we separated the trap component of each device. Also, we calculated the total trap density using the conventional charge pump measurement method. [4-5]

In addition, TEG can be used to predict global reset operation. As illustrated in figure 5, 3 chip-level data with different process condition shows different amount of overflowing charges from PD to SG. It is confirmed that the S/O voltage of OG is lowered in the TEG data under the process condition where the amount of overflow charge increased. This shows that the TEG and the chip-level results are correlated.

We correlated the chip-level dark current and TEG result. First we designed devices with different SG sizes, and compared the number of traps calculated from TEG charge pump measurement and chip-level dark current. Figure 6 (a) illustrates the correlation between chip-level dark current and the number of traps calculated from

TEG. Next, maximum capacitance (Cmax) measured from TEG also be correlated with dark current in chip-level in different process condition. As shown in figure 6 (b), we confirmed that the number of border traps is increased through TEG C-V measurement, and is related to the increase of chip-level dark current.

References

[1] Taesub Jung, A 4-tap global shutter pixel with enhanced IR sensitivity for VGA time-of-flight CMOS image sensors, Electronic Imaging (2020)

[2] Rahman, M et al., Border trap extraction with capacitance-equivalent thickness to reflect the quantum mechanical effect on atomic layer deposition high-k/$In_{0.53}Ga_{0.47}As$ on 300-mm Si substrate. Scientific Reports, 9(1), 1-12.

[3] Misra, D. (2011). High k dielectrics on high-mobility substrates: The interface!, The Ele. Chem. Sc. Interface, 20(4), 47.

[4] Jungwook Lim, Development of Test Element Group Methodology for Time-of-Flight CMOS Image Sensor, International Conference on Solid State Devices and Materials (2022)

[5] Groeseneken G., R. F. (1984). A reliable approach to charge-pumping measurements in MOS transistors. IEEE TED, 31(1), 42-53.
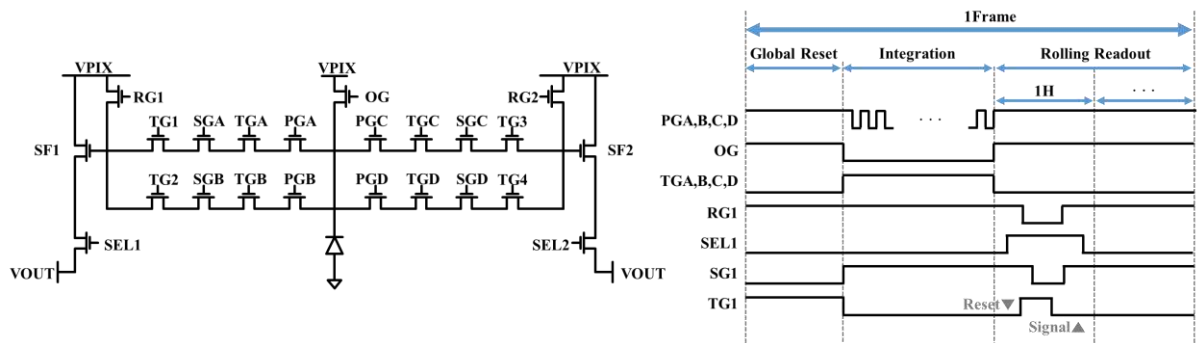
Figure 1. Schematics and timing diagram of 4-tap iToF unit pixel
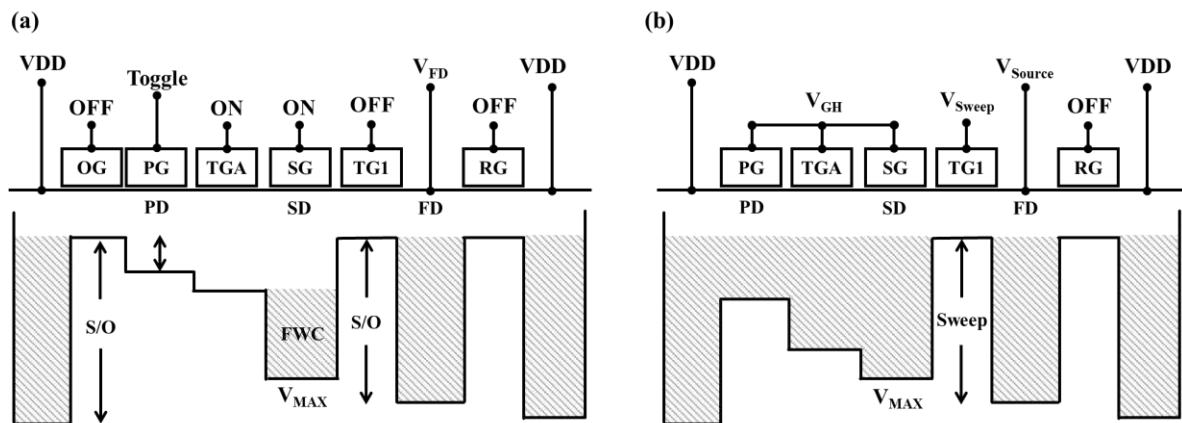
Figure 2. (a) Potential diagram of iToF in modulation (1 of 4 taps is shown)
(b) Measurement setup for TG S/O voltage in TEG device

Figure 3. (a) Measured electrical potential of TEG device (b) chip-level FWC data with different implant condition

Figure 4. (a) C-V curve measurement of SG (b) the equivalent circuit (c) band diagram of interface trap and border trap

Figure 5. (a) Charge overflow from PD to SG measured from chip-level (b) TEG measurement result of shut-off voltage of OG at the different process condition (c) Potential diagram of the charge overflow from PD to SG

Figure 6. (a) The correlation between chip-level dark current and the number of traps calculated from TEG with different SG size (b) The correlation between chip-level dark current and the maximum TEG capacitance

Silicon Metalenses towards a fully Silicon integrated SWIR sensing

Matthieu Dupre[1], Jian Ma[1], Biay-Cheng Hseih[1] and Sergio Goma[1],
[1] QCT Multimedia R&D – Imaging, Qualcomm Technologies Inc.,
5775 Morehouse Drive, San Diego CA 92121 USA
mdupr@qti.qualcomm.com

Most Silicon based depth and lidar sensors rely on near-infrared (NIR 750-900nm) sources to produce depth images as Silicon CMOS sensors can achieve a high quantum efficiency for an unbeatable cost at such wavelengths [1]. Some automotive LIDAR is using 900nm to 1400nm illuminations with InGaAs sensors with small advantages compared to NIR [2]. Advances in short wave infrared (SWIR) sensor technologies, such as Silicon-Germanium sensors [3], changes this paradigm and opens a new window for groundbreaking sensor designs, as SWIR can push the wavelength above retinal hazard area (>1400nm), allowing for much higher eye safety, due to the low penetration of those wavelengths through the eye lens (IEC 60825-1 Edition 3.0 2014-05, ANSI Standard Z136.1)[4]. Here, we propose to use a Silicon Metalens flat optics and build upon our stacked sensor technologies [5] to obtain a fully Silicon integrated stacked sensor at SWIR wavelengths. We will discuss the design of the stacked sensor and focus on the Silicon Metalens for multiple use-cases. We will show numerical simulations of the optical stack for eye-tracking application or wide-angle time of flight (TOF) and how we can obtain very compact form factor modules. Finally, we will demonstrate the results of our Silicon metalens prototype at 1550nm.

SWIR wavelengths are specifically interesting for sensing technologies such as TOF or eye-tracking. Firstly, the solar spectrum exhibits a dip around 1380nm: there is no solar background illumination, which leads to less shot-noise on the sensor and much higher signal to noise ratio for outdoor applications. In addition, SWIR wavelengths are much safer for the eye, opening new possibilities for eye-tracking whether it is for extended reality application or driver monitoring in automotive. Finally, at SWIR wavelengths Silicon becomes completely transparent with a high index of refraction (n=3.5), allowing for a simpler manufacturing process of BSI sensors, as back thinning is no longer required. Here, we leverage this property by designing a lens made of a single Si wafer with the metalens technology.

Figure 1 shows our proposed stacked sensor. It is made of an optional aperture wafer (glass or other low index material) and an eventual bandpass filter layer. Interestingly, glass with an index around 1.5 can play the role of a low index material in this case, as Silicon has a much higher optical index, close to 3.5 at SWIR wavelengths. Hence, we can use a first spacer made of glass or any other low index material. Using a simple Silicon spacer is also possible at the cost of a thicker design but allows easier wafer bonding between the different layers. The second layer is a Silicon Metalens, directly patterned on a Silicon wafer. Then, we use another spacer: it can be a glass wafer if we need a compact low index material, a Silicon Spacer wafer, or simply an unpolished backside illuminated sensor. Photons are collected on a SWIR sensor such as a SiGe sensor. The latter is directly placed upon a ROIC ASIC wafer on top of a reconfigurable chip, such as a Reconfigurable Instruction Cell Array (RICA) [5]. The advantage of processing at the edge is significant for low-power applications [6] and programmability can enable a broad range of image processing algorithms fast and efficiently. The multiple wafers are stacked with wafer-bonded technology to get a fully monolithic Silicon camera.

As examples of possible applications, we discuss a 2.5 µm pixel pitch eye-tracker sensor for mixed-reality glasses or headset and an iTOF sensor with 5 µm pixel pitch. The latter needs a larger pixel pitch to accommodate several taps for the several depth sensing channels. SiGe sensors with such pixel pitches are easily obtainable with current technologies. Figure 2 shows Zemax ray-tracing simulations of the optical stacks of such an eye-tracker (a,b and c) and a iTOF sensor (d, e, and f). Fig.2a) and d) show the layout of the devices for incident field angle of 0, 10, 20, 30, 40 and 45 degrees. Fig.2.b) and e) show the spot diagrams for the two systems which are close to the diffraction limit. Finally, Fig.2c) and f) show the MTF of the optical systems up to their Nyquist frequency. The two examples exhibit high MTF of 55 % at Nyquist/2 for the eye-tracker over 70% for the iTOF sensor. Table 1 summarizes the different key parameters of a 400x400 pixels eye-tracker and a 1920x1080 pixels iTOF sensor with a silicon metalens.

The metalens phase law can be optimized using Ray-tracing software such as Zemax [7]. As a conventional lens can be optimized using aspherical terms, we propose to model it by the sum of the hyperbolic phase-law and aspherical terms:

1

$$OPD(r) = a_0\left(\sqrt{r^2 + f^2} - f\right) + \sum_{m=2}^{10} a_m r^m \qquad (1)$$

where r is the radial position on the metalens normalized by the metalens radius and f the focal length. a_0 is a term controlling the ratio of the hyperbolic phase law with respect to the aspherical terms. It can also be used to conveniently adapt to different material. Basically, a hyperbolic metalens would have a_0 equals to the index of the material. For instance, to design a metalens focusing a normally incident beam into air, $a_0=1$, into glass $a_0=1.5$ and into Silicon $a_0=3.5$. Its sign becomes negative to model diverging lenses.

Figure 3 shows our Metalens prototype. It is made in a 500 μm thick Silicon wafer in which are etched 1200nm-long nanopillars. Those nanopillars are placed along a hexagonal unit cell with a lattice constant of 500nm. An electronic microscope image of our metalens is Shown on Fig.3a). A picture of the Metalens is shown in Fig.3b). By locally controlling the nanopillars diameters (varying between 100nm and 400nm), we can tune the phase shift the light undergoes while traversing the structure to realize a lens. Figure 3c) shows an image of a USAF target taken with our metalens prototype and an off-the shelf SWIR sensor (InGaAs) at a 1550 nm wavelength.

Those first results prove the feasibility of a fully silicon manufacturable camera based on a SWIR metalens, opening the way for a new generation of programmable smart cameras, from eye-tracking for mixed reality or driver monitoring to depth sensing for automotive or security applications. Our next step consists in switching from a R&D prototype made with e-beam lithography to an industrial process involving UV lithography to obtain a complete monolithic silicon camera fully manufacturable in a regular CMOS fab that requires no further packaging.

[1] E. Marti, M. A. de Miguel, F. Garcia and J. Perez, "A Review of Sensor Technologies for Perception in Automated Driving",*IEEE Intelligent Transportation Systems Magazine, 2019,14, 4,*94-108..

[2] Nicolas Pinchon et al., et al. "All weather vision for automotive safety: which spectral band?", *AMAA 2018, Advanced Microsystems for Automotive Applications*, 2018, 3-15.

[3] N. Na *et al.*, "High-Performance Germanium-an-Silicon Lock-in Pixels for Indirect Time-of-Flight Applications," *2018 IEEE International Electron Devices Meeting (IEDM)*, 2018, 32.4.1-32.4.4.

[4] https://ohsonline.com/articles/2014/08/01/laser-safety_0.aspx

[5] B. C. Hseih et al., "A 3D Stacked Programmable Image Processing Engine in a 40nm Logic Process with a Detector Array in a 45nm CMOS Image Sensor Technologies", *International Image Sensor Workshop*, 2017.

[6] Jorge Gomez and al., "Distributed On-Sensor Compute System for AR/VR Devices: A Semi-Analytical Simulation Framework for Power Estimation", *2022 tinyML Research Symposium*.

[7] A. Arbabi et al. "Miniature optical planar camera based on a wide-angle metasurface doublet corrected for monochromatic aberrations", *Nature Communications*, 2016, 7.

[8] M Shalaginov et al. "Single-Element Diffraction-Limited Fisheye Metalens", Nano Letters, 2020, 20, 10, 7429-7437.

[9] Engelberg J, et al. "Near-IR wide-field-of-view Huygens metalens for outdoor imaging applications", *Nanophotonics,* 2020, 9, 361–370.

Aperture Wafer
SWIR NB Filter
Si or Quartz Spacer Wafer
Metalens Si Wafer
Si or Quartz Spacer Wafer
SWIR Detector SiGe Wafer
ROIC ASIC Wafer
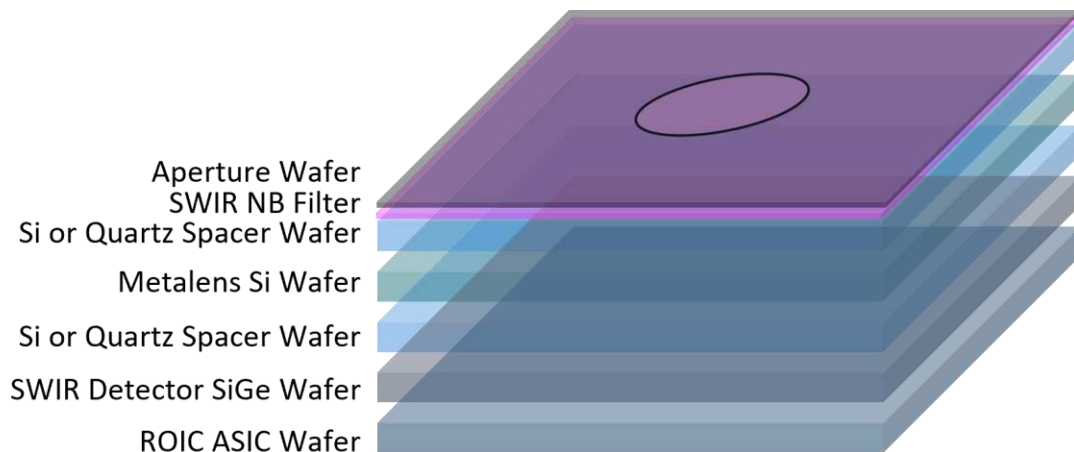
Figure 1. Stacked Silicon camera sketch.
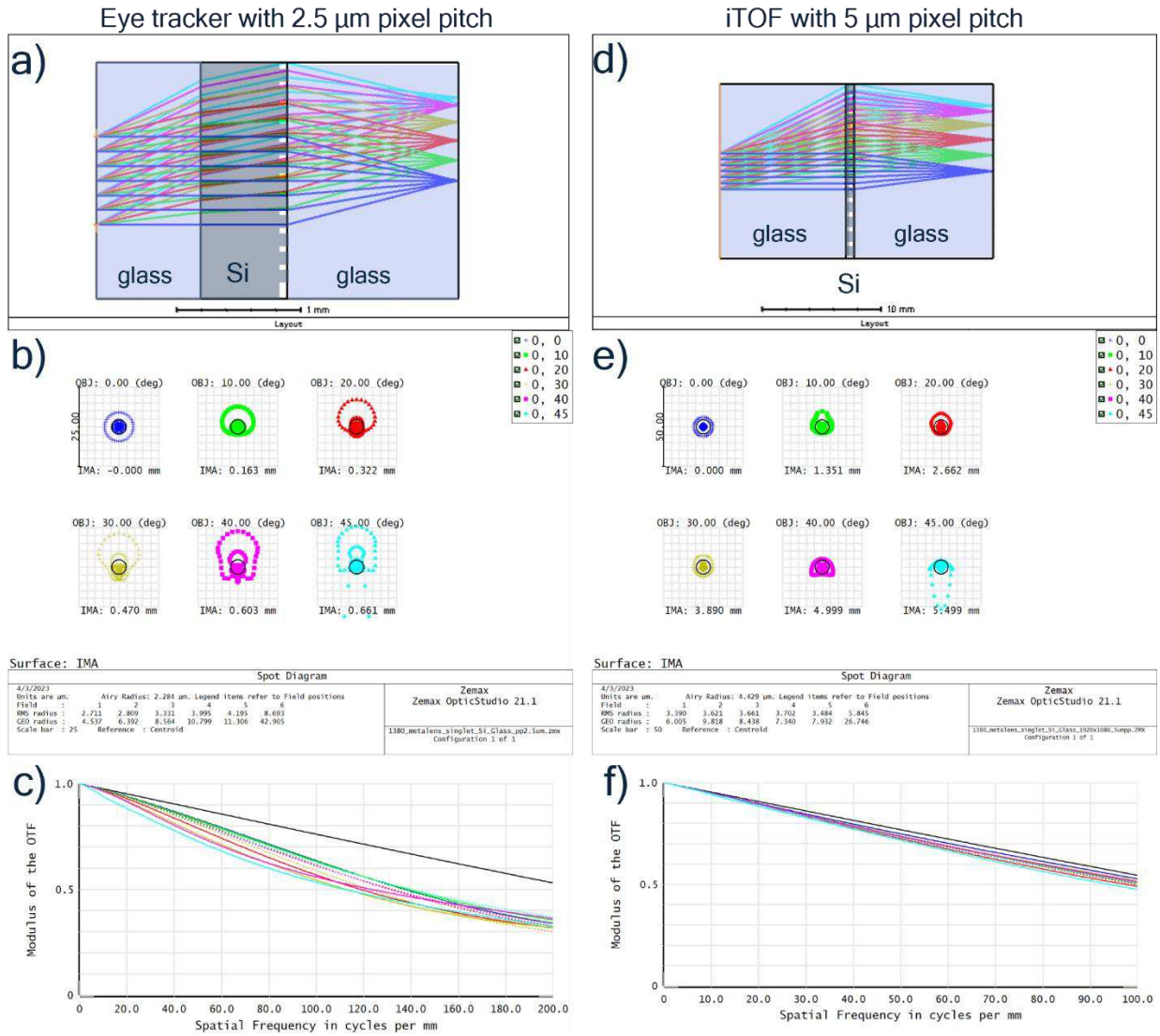
Figure 2. Ray tracing (Zemax) simulations of an eye-tracker and iTOF sensors.

	Eye-tracker	iTOF
Resolution	400x400	1920x1080
Pixel pitch	2.5 µm	5 µm
z height	2.9 mm	23 mm
FOV	90°	90°
EFL	0.9 mm	7.65 mm
Aperture Ø	0.35 mm	3 mm
F#	2.6	2.55
MTF @ Ny/2	55%	70%

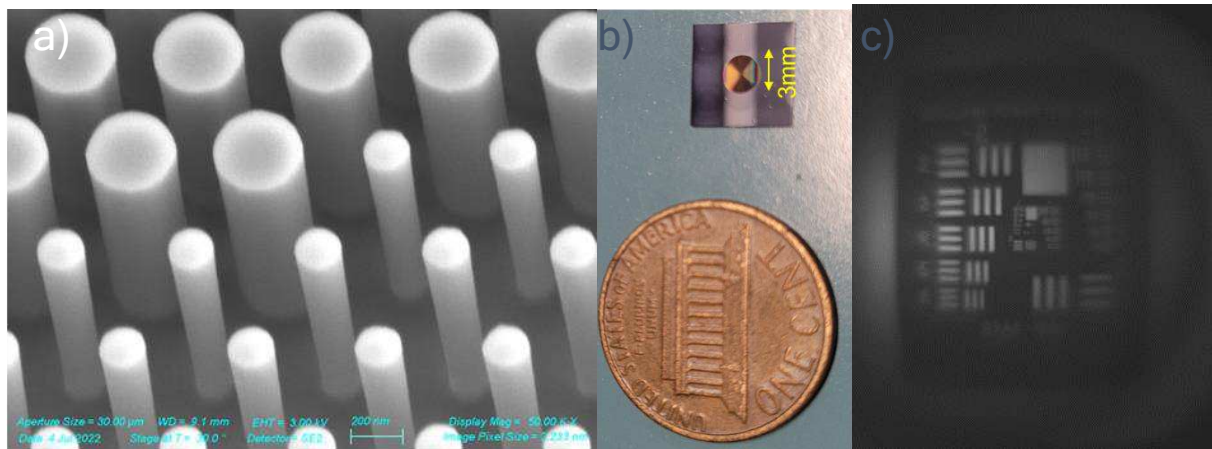*Table 1. KPI comparisons of two SWIR sensors for eye-tracking and indirect Time Of Flight (iTOF).*

Figure 3. a) Electronic microscope image of zoom of the 1550nm Silicon metalens. b) Picture of the 3mm diameter metalens. c) Image recordeed at 1550nm using a SWIR InGaAs sensor.

4

Toward a Photon Counting Detector for X-ray Imaging by Direct Deposition of Scintillator on 32x32 CMOS SPAD Array

Yao-Lun Liu[1], Meng-Hsuan Lu[1], Chun-Hsien Liu[1], Yu-Hsun Wu[1], Sheng-Di Lin[1], Chia-Ming Tsai[1], Jin-Cherng Hsu[2], Yu-Cheng Syu[3], and Jau-Yang Wu[4*]

[1]Institute of Electronics, National Yang Ming Chiao Tung University, 1001 University Road, Hsinchu 300, Taiwan

[2]Graduate Institute of Applied Science and Engineering, Fu Jen Catholic University, New Taipei City 242062, Taiwan

[3]Department of Physics, Fu Jen Catholic University, New Taipei City 242062, Taiwan

[4] Department of Electrical Engineering, Yuan Ze University, Taoyuan 32003, Taiwan

*Corresponding Author: judewu13@saturn.yzu.edu.tw , telephone: +886-3-4638800 ext. 7520

Abstract—We develop a low cost and high performance X-ray imager by directly depositing the CsI material on the 32x32 CMOS single photon avalanche diode (SPAD) array as a scintillator. From the effective dose calculation, our imager performs more sensitive compare to a photomultiplier tube (PMT) with scintillator which is often used in the X-ray diffractometer (XRD). So we believe that the SPAD array imager with direct deposition of CsI could be a better candidate for digital radiography systems.

I. INTRODUCTION

Combining scintillator materials with silicon based photon detector can make up a low cost digital radiography image system as a replacement of the traditionally radiographic films. The scintillator can convert the incident X-rays to optical light in the detection wavelengths range (420-600 nm) of silicon photon detectors. However, the indirect detection method has low detection efficiency because of the low transfer efficiency of the scintillator materials [1], leading to long integration time and poor spatial resolution in the applications such as dentistry, surgery, positron emission tomography (PET) [2], and mammography [3]. The CMOS SPAD array is a kind of imager which possess ultra-high photon sensitivity and easy to be integrated with digital signal processing circuits, so a scintillator deposited on COMS SPAD array can realize digital radiography systems with high frame rate and low cost [4].

II. THE DEVICE FABRICATION AND TESTING METHODS

For the traditional method that uses crystalline (slab) scintillators coupled to the imager, the light transferred from scintillator will experience multiple reflections between several interfaces, causing laterally spread of light and resulting in poor spatial resolution and lower optical efficiency [5]. In this work, differing from the conventional method, we deposit an undoped cesium iodide (CsI) films as the scintillator directly on a 32x32 SPADs array to perform better spatial resolution and photon coupling efficiency [6]. The CsI film is prepared by thermal vacuum evaporation technique. The cross sectional structure of SPAD array with CsI coating is shown in Fig. 1(a), and the top view picture of final chip is shown in Fig. 1(b). From the cross-sectional image of dual-beam focused ion beam (FIB) microscope as shown by Fig 2(a), the coating of CsI has the thickness of
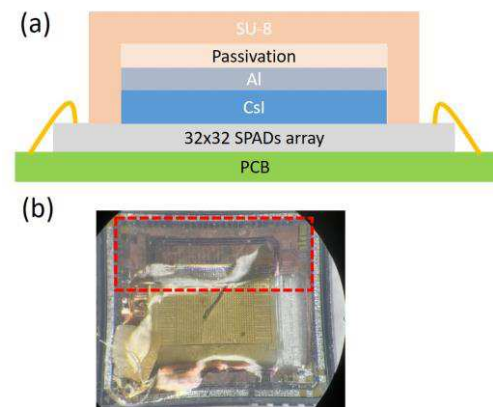
Fig. 1. Cross-section structure of 32x32 SPAD array chip with directly deposited CsI layer (a) and the top view picture of chip. The red dotted region indicates the area without CsI coating.

about 14 µm. The image shown in Fig. 2 (b) clearly shows that the microcolumn width is about 0.5 µm for the deposited CsI of microcolumns array type. A columnar-structure with needle-like tips is also observed for CsI films. The size and shape of deposited CsI depends on the temperature and
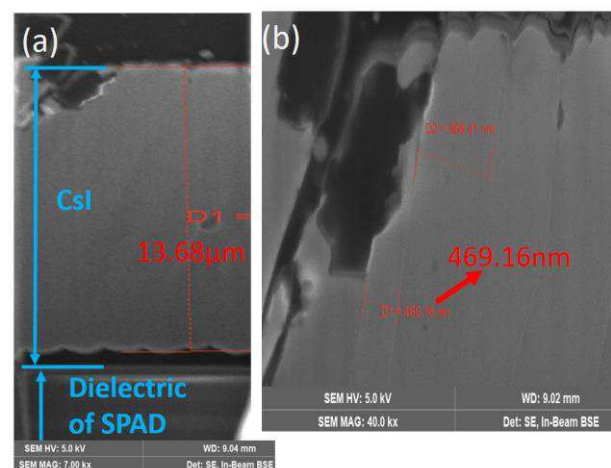
Fig. 2 (a) Cross-sectional picture of CsI from focused ion beam microscope. (b) Magnification of cross-sectional picture.

pressure during the thermal vacuum evaporation. Note that the cross-sectional FIB picture shown here is taken from the device different with the one under investigation. The passivation and SU-8 layer are used for preventing CsI layer from hydrolysis. For examining the effect of CsI layer, we purposely designed a region without CsI coating as shown by the red dotted frame in Fig. 1(b) [7, 8]. The 32x32 SPAD array is fabricated by TSMC 0.18 um BCD HV process, with a breakdown voltage of 48.5 V. The SPAD array is integrated with time gated quenching circuit [9].

The chip readout is controlled by a FPGA (kintex-705). The dark count rates (DCRs) of the chip are almost unaltered after the CsI coating, where it is plotted in Fig. 3(a). There is an aluminum film upon the CsI coating for blocking the light other than the emission from CsI layer. Fig. 3(b) shows that there is higher photon counts of SPADs for the area without CsI coating when the chip is under illuminated.

Fig. 4. Picture of measurement system setup in the X-ray diffractometer (XRD).

Fig. 3 (a) Measured dark count rates of 32x32 pixels (b) and the photon count rate under ambient light.

The measurement system setup in the X-ray diffractometer (XRD) instrument is shown in Fig. 4. First a lower SPADs bias voltage (51.4 V) was used for the preliminary test. The DCRs of chip before and after X-ray irradiation has no obvious difference, so the chip was not damaged by the irradiation. Fig. 5 shows that there is obvious photon counts of SPADs with the CsI coating when X-ray gun was turned on, where the photon counts were measured with integration time of 0.1 second. A higher bias voltage (52.5 V) is used to achieve higher photon detection efficiency of SPADs for the following measurements, acquiring much higher photon

Fig. 5. The dark count rate of SPAD array under the bias voltage of 51.4 V and 0.1 second integration time (a) before X-ray irradiation, (b) under X-ray irradiation and (c) after X-ray irradiation
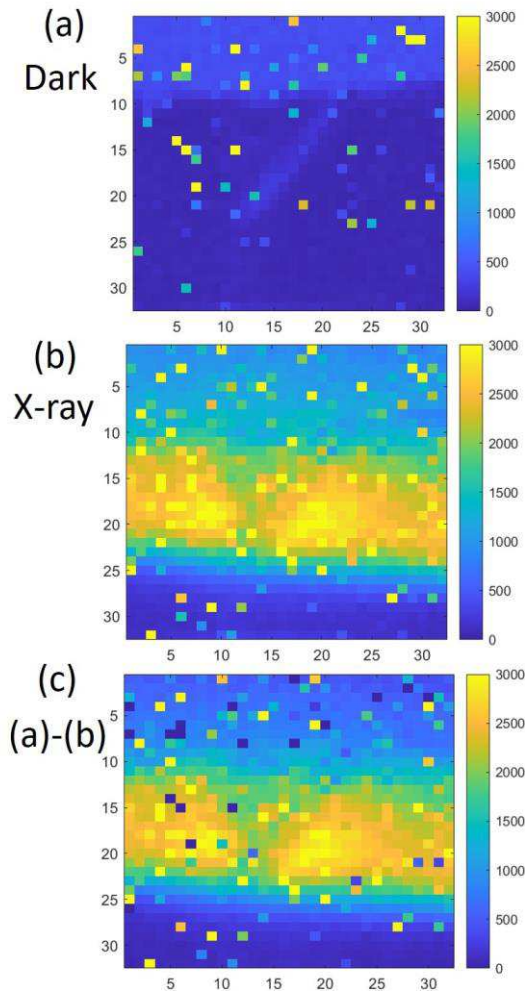
Fig. 6. The dark count rate of SPAD array under the bias voltage of 52.5 V and 0.1 second integration time (a) before X-ray irradiation and (b) under X-ray irradiation. (c) The net photon counts with counts of (b) subtract from the counts of (a).
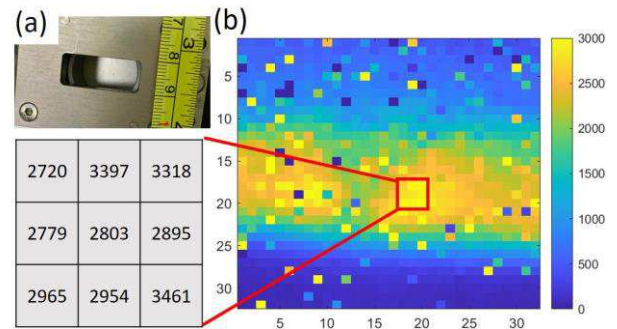

Fig. 7. (a) Picture of the PMT detector in XRD with an opened window of 2.5 x 1 cm^2. (b) Photon counts of 9 pixels of our SPAD array, measured in 0.1 s integration time.
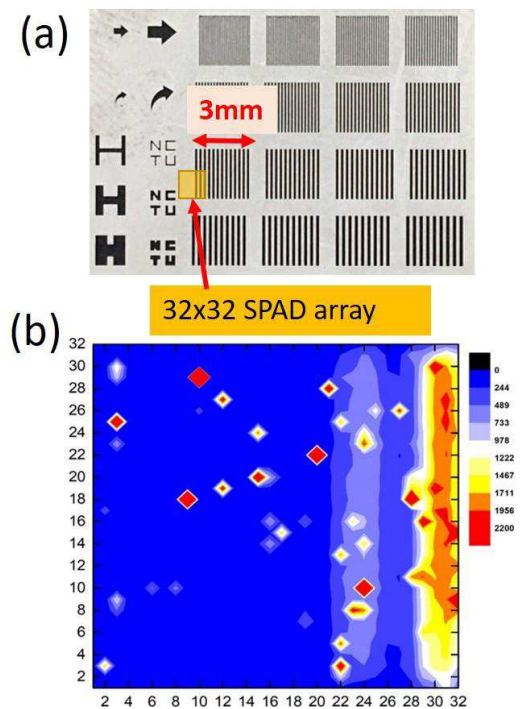

Fig. 8. The picture of stainless shutter with multi-slit patterns and the position of 32x32 SPAD arrays under the multi-slit of 130 um width (a), the radiography of the measurement in XRD.

counts in the X-ray irradiation measurement as shown in Fig. 6. The highest net photon counts are over 3000 counts during the integration time of 0.1 second as shown in Fig. 6 (c).

To examine the effect of CsI film on the performance of our chip for the X ray detection measurement, we replace the detector by our chip to perform X ray detection. Then we compare the effective photon counts of our chip to that of the PMT detector with scintillator which is used in the original XRD system. The PMT detector with an opened window of 2.5 x 1 cm^2, as shown in Fig. 7(a), can detect about 600 counts per second (cps) with a detection area of 20-µm diameter, where we normalize the photon counts to the area size same as that of our device for better comparison. As shown in Fig. 7 (b), under the same condition, our chip can detect over 30000 cps averaged from 9 pixels of our chip. The photon counts detected by PMT is much lower than the that detected by our chip.

To demonstrate a digital radiography system, we use a stainless shutter with various multi-silts width pattern ranging

from 50 µm to 200 µm, as shown in Fig. 8 (a), and put the shutters in between the chip and X-ray gun. The 130 µm width of multi-slit is used for radiography. Although the stainless shutter was not aligned well with the SPAD, the

multi-slit. radiography can still be clearly measured as shown in the Fig. 8 (b). The digital radiography picture shows good spatial resolution due to the benefit of direct deposition of scintillator on SPAD array.

III. Conclusion and Discussion

From the measurement results and effective dose calculation, our imager performs much higher sensitive detection comparing to a photomultiplier tube (PMT) with scintillator which is used in the X-ray diffractometer (XRD) , thanks to the highly photon sensitive of 32x32 pixels CMOS SPAD array. We also demonstrate a high spatial resolution radiography with our imager, verifying the benefit of directly depositing the CsI material as the scintillator on the SPAD imager. So we believe that the SPAD array imager with direct deposition of CsI could be a better candidate for digital radiography systems.

Furthermore, in order to develop a digital radiography imager with timing information, the luminescence lifetime information of our deposit CsI is an ongoing task. Our chip with high speed operation is suitable for the high X-ray dose application, though the X-ray induced damage of scintillator should be prevented [10].

IV. Acknowledgment

V. References

[1] Wang, Wenzhen, et al. "Approaching the theoretical light yield limit in CsI (Tl) scintillator single crystals by a low-temperature solution method." Crystal Growth & Design 20.5 (2020): 3474-3481.

[2] Xu, Chen, et al. "Comparison of digital and analog silicon photomultiplier for positron emission tomography application." 2013 IEEE Nuclear Science Symposium and Medical Imaging Conference (2013 NSS/MIC). IEEE, 2013.

[3] Patt, B. E., et al. "High resolution CsI (Tl)/Si-PIN detector development for breast imaging." IEEE Transactions on Nuclear Science 45.4 (1998): 2126-2131.

[4] Prekas, G., et al. "Direct Deposition of Microcolumnar Scintillator on CMOS SSPM Array: Toward a Photon Counting Detector for X‐Ray/Gamma Ray Imaging." AIP Conference Proceedings. Vol. 1412. No. 1. American Institute of Physics, 2011.

[5] Venialgo, Esteban, et al. "Time estimation with multichannel digital silicon photomultipliers." Physics in Medicine & Biology 60.6 (2015): 2435.

[6] Hsu, Jin-Cherng, and Yu-Shen Ma. "Luminescence of CsI and CsI: Na films under LED and X-ray excitation." Coatings 9.11 (2019): 751.

[7] Nagarkar, V. V., et al. "Structured CsI (Tl) scintillators for X-ray imaging applications." IEEE transactions on nuclear science 45.3 (1998): 492-496.

[8] Nikl, Martin. "Scintillation detectors for x-rays." Measurement Science and Technology 17.4 (2006): R37.

[9] C. C. Hsu, et al. "CMOS Single-photon Avalanche Diodes using Gated Reset Circuit with On-chip Pulse Width Modulation," P09, IISW 2019.

[10] Tremsin, A. S., et al. "X-ray-induced radiation damage in CsI, Gadox, Y2O2S and Y2O3 thin films." Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment 459.3 (2001): 543-551.

Feedback Control of a Block-Wise-Controlled Image Sensor Based on Brightness Distribution Analysis

Kohei Tomioka[1], Kodai Kikuchi[1], Takenobu Usui[1], Kazuya Kitamura[1], Shoji Kawahito[2]

[1] NHK Science & Technology Research Laboratories, Tokyo, Japan
[2] Research Institute of Electronics, Shizuoka University, Hamamatsu, Japan
TEL: +81-3-5494-3326 E-mail: tomioka.k-dk@nhk.or.jp

Abstract This study proposes an image sensor design implementing block-wise control that allows independent control of pixel binning and exposure time within each pixel block. This facilitates the flexible control of the frame rate, resolution, and dynamic range for better applicability to a shooting scene. A 1 K × 1 K prototype image sensor with 16 × 17 blocks (64 × 64 pixels in each block) is used to demonstrate the feedback control capability of the image sensor based on brightness distribution analysis.

I. INTRODUCTION

Tradeoffs between frame rate, resolution, noise performance, and dynamic range render the designing of high-pixel-rate image sensors [1] required for recent imaging systems such as 8 K [2], VR, and 360° video [3] difficult. To overcome these tradeoffs, this study proposed a scene-adaptive imaging system that facilitates local control of the imaging parameters based on individual areas according to the characteristics of the scene, such as object movement and brightness distribution. Therefore, imaging parameters such as the resolution and frame rate can be allocated according to the scene being recorded. In addition, since the exposure time can be controlled for each area, it is possible to increase the exposure time in dark areas to improve S/N, and shorten the exposure time in bright areas to expand the dynamic range. A key technology in this system is the proposed architecture for block-wise-controlled image sensors. This architecture divided the pixel array into blocks, wherein an external feedback signal is used to individually control pixel binning and exposure time. This feedback signal was obtained based on an external scene analysis and specified the appropriate imaging parameters for each area. This study demonstrated the feedback control of a block-wise-controlled image sensor, wherein the operation modes could be changed locally based on brightness distribution analysis.

II. BLOCK-WISE-CONTROLLED IMAGE SENSOR

Figure 1 shows a block diagram of the proposed system, which comprised a block-wise-controlled image sensor and a signal processor. The signal processor performs a brightness distribution analysis of the captured scene and determined the optimal operating mode for each block based on its brightness. Consequently, the result is fed back to the image sensor as a feedback signal. The image sensor operates under different modes for each block according to the feedback signal. In this system, analysis processing and feedback operations is performed within one frame period to facilitate real-time response to changes in the scene.

Figures 2 and 3 show the die image and pixel architecture of the image sensor, respectively. The prototype image sensor comprised a 1024 × 1088 pixel array, pixel driver, column-parallel ADCs, an output block, and a mode controller. The die dimensions are 6.5 mm (H) × 7.7 mm (V). A 1024 × 1088 pixel array is divided into 16 × 17 control blocks (64 × 64 pixels per block). Figure 3 (c) shows that every 2 × 2 pixel (depicted as *A*, *B*, *C*, and *D*) shares a pixel amplifier and receives readout pulses through the switches provided for each pixel. Therefore, a selected pixel or a pixel-binned signal of 2 × 2 pixels can be selectively read according to the control signal. These switches are independently controlled by the control signals specified by the mode controller for each block according to the externally received feedback signals. Further, the output block output the sensor data as an LVDS 4-ch signal.

Table 1 summarizes the four operational modes supported by the image sensor, and Fig. 4 shows the readout scanning method for these modes. Compared to the results presented in Ref. [4,5], the proposed image sensor facilitates control of the exposure time and frame rate for each area. The scanning method is as follows. (a) Normal mode: One selected pixel signal is read for each scan, thus enabling subframe readouts with 1/240 s periods in the order of A, B, C, and D. The exposure time is 1/60 s, and the resolution is 64 × 64 pixels per block. (b) Fast mode: Pixel-binning readout is performed for each scan. This enables high-speed readouts with a frame rate of 240 fps. However, the resolution deteriorates to 1/4 (32 × 32 pixels per block). (c) Bright mode: Sub-frame readout is

performed in a manner similar to that in the normal mode; however, the exposure time is limited to 1/240 s using an electronic shutter. (d) Low-light mode: Sub-frame readout and 4-frame readout pauses are alternately performed, thereby extending the exposure time to 1/30 s.

III. BRIGHTNESS DISTRIBUTION ANALYSIS

The brightness distribution analysis and feedback signal generation processes are shown in Fig. 5. First, the signal of subframe A is averaged every 8×8 pixels, where averaged signal number in one block is 64. The averaged results is classified into low-light, normal, and bright modes using two thresholds (depicted as thresholds 1 and 2 in Fig. 5). Finally, the mode with the highest frequency of each of the three modes is determined for each of the 16×17 blocks using a mode filter. These operation modes are fed back to the image sensor as feedback signals. The output signal level varies depending on the mode. In particular, in low-light mode, the signal level is twice as high as normal mode; and in bright mode, it is 1/4 of the signal level of normal mode. To correct this signal-level difference in signal processor, the level correction factors ($\times 1/2$, $\times 1$, $\times 4$) are multiplied by the next subframe A, depending on each operating mode. These processes are implemented on an FPGA board. Further, processing for analysis and feedback signal generation is performed within 1/60 s. Therefore, each block of the image sensor can be operated in real-time in the optimal operating mode according to changes in the brightness distribution of the scene.

IV. EXPERIMENTAL RESULTS

To verify the block-wise control function, the operation modes are set using a specified external feedback signal. Fig. 6 shows images of the four operation modes set by the specified external feedback signal. The images in each mode correspond to their imaging parameters. In the fast mode, pixel binning results in an image with 1/4 resolution compared to the normal mode. In the bright mode, the signal value is 1/4 because of the exposure control of 1/240 s. Finally, in the low-light mode, the signal value is twice that in normal mode because the exposure time is extended to 1/30 s.

Figure 7 shows the video images and feedback signals captured by the proposed system. The upper row of the four images shows the captured video images, whereas the lower row of four images shows the color-coded results of the brightness distribution analysis. Depending on the brightness distribution, the three different modes: low-light, normal, and bright, are assigned. The area captured in low-light mode is changed to follow the area of the moving hand, whereas that where the LED light is captured is switched to high-brightness mode. Further, the region captured by the transmitted chart, which exhibits intermediate brightness, switches to normal mode. Thus, these results demonstrate the feedback control capability of the operation modes based on brightness distribution analysis.

V. CONCLUSION

This study proposes a scene-adaptive imaging system that facilitated local control of the imaging parameters based on individual areas according to the characteristics of the scene. The prototype image sensor demonstrates block-wise control of the resolution, frame rate, and exposure time using the specified external feedback signal. In addition, the proposed system demonstrates the feedback control capability of the operation modes based on a brightness distribution analysis. Thus, the results of this study indicate that the architecture of the prototype image sensor and the proposed system are suitable for realizing scene-adaptive imaging.

V. REFERENCES

[1] S. Kawahito, "Column-Parallel ADCs for CMOS Image Sensors and Their FoM-Based Evaluations," IEICE Trans. Electron., vol. E101-C, no. 7, pp. 444–458, 2018.

[2] Recommendation ITU-R BT. 2020: "Parameter Values for Ultra-High Definition Television Systems for Production and International Programme Exchange," 2015.

[3] Recommendation ITU-R BT. 2123: "Video Parameter Values for Advanced Immersive Audio-Visual Systems for Production and International Programme Exchange in Broadcasting," 2019.

[4] T. Hirata et al., "A 1-inch 17 Mpixel 1000 fps Block-Controlled Coded-Exposure Back-Illuminated Stacked CMOS Image Sensor for Computational Imaging and Adaptive Dynamic Range Control," International Solid-State Circuits Conference (ISSCC), pp. 120–121, 2021.
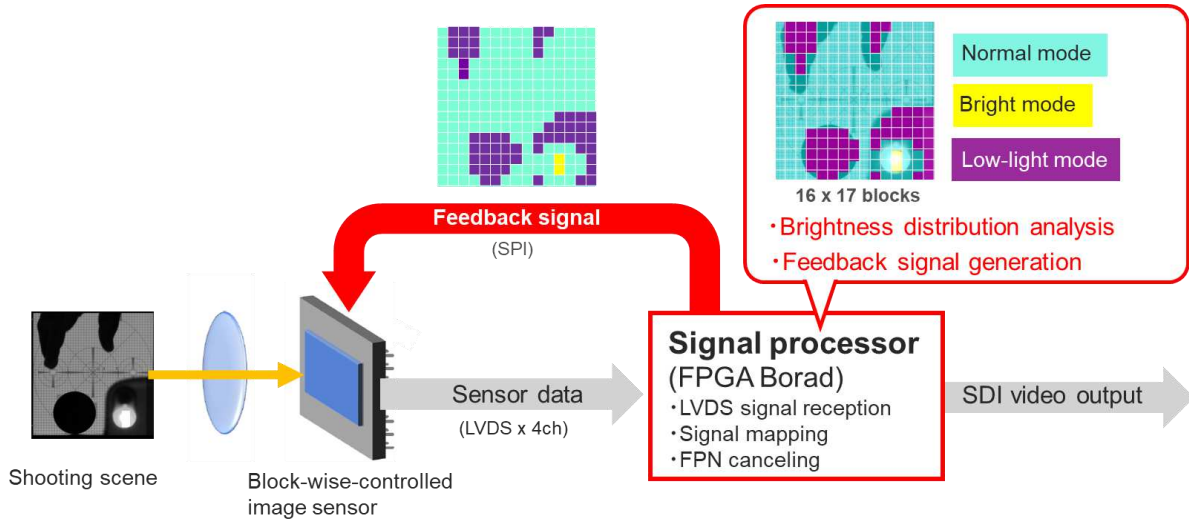
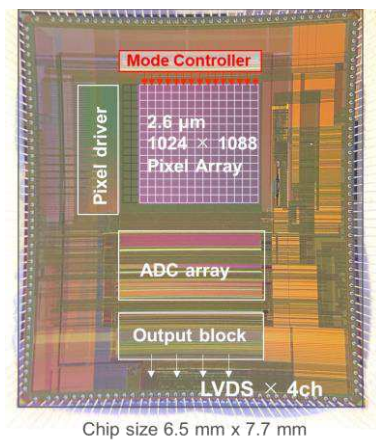Fig. 1. Block diagram of the prototype scene-adaptive imaging system.
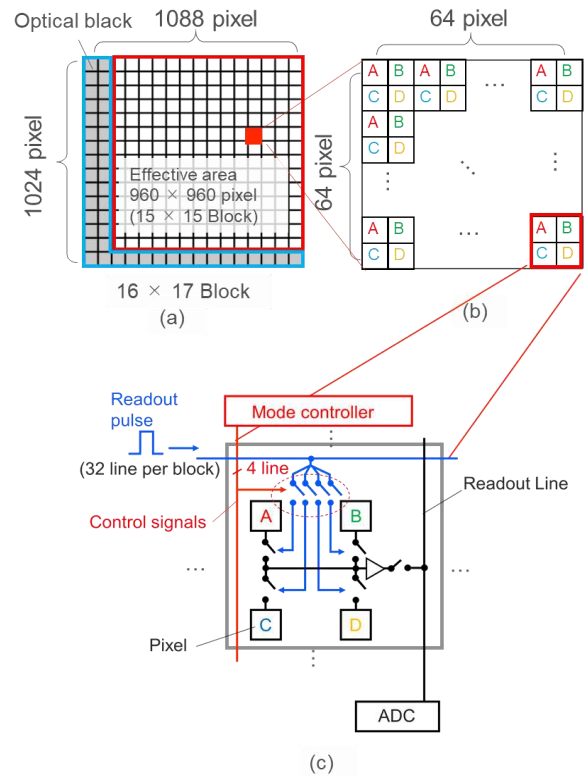

Fig. 2. Die image of the image sensor.

Table 1. Operation modes implemented in the image sensor.

Mode	Resolution	Frame rate	Exposure time
Normal	64 × 64	60 fps	1/60 s
Fast	32 × 32	240 fps	1/240 s
Bright	64 × 64	60 fps	1/240 s

Fig. 3. Schematic of the pixel architecture: (a) pixel array, (b) pixel block, and (c) pixel readout circuit for 2 × 2 pixels.
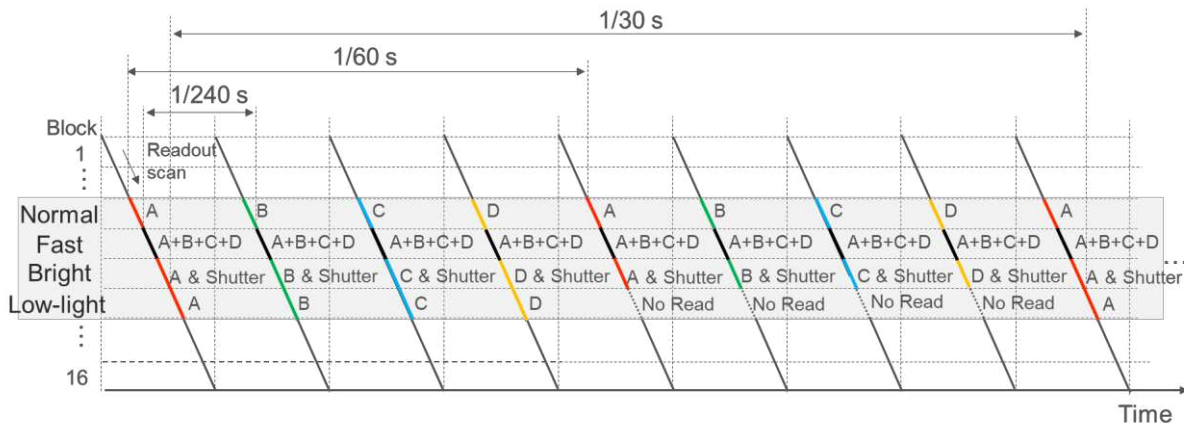
Fig. 4. Schematic of the readout scan method of the image sensor.
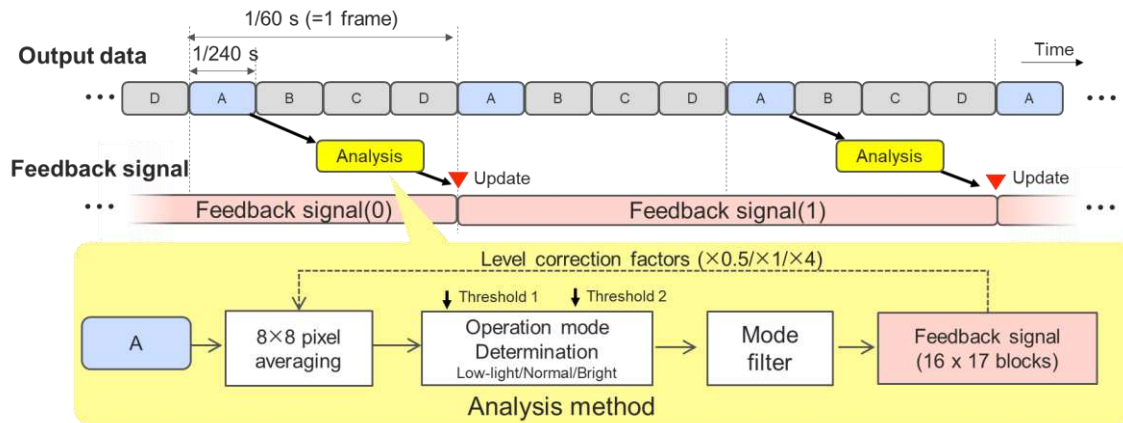
Fig. 5. Schematic of the brightness distribution analysis and feedback signal generation processes.
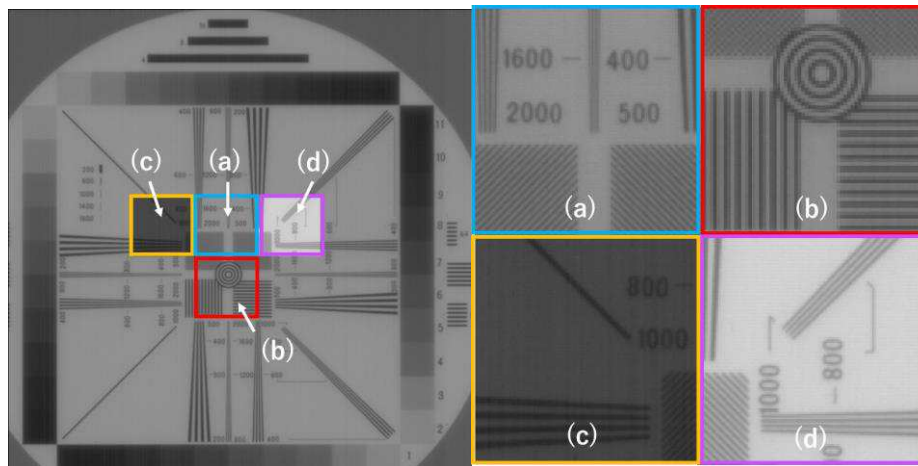
Fig. 6. Captured images by the prototype scene-adaptive imaging system with the operation mode set by external feedback signals: (a) normal, (b) fast, (c) bright, and (d) low-light modes.
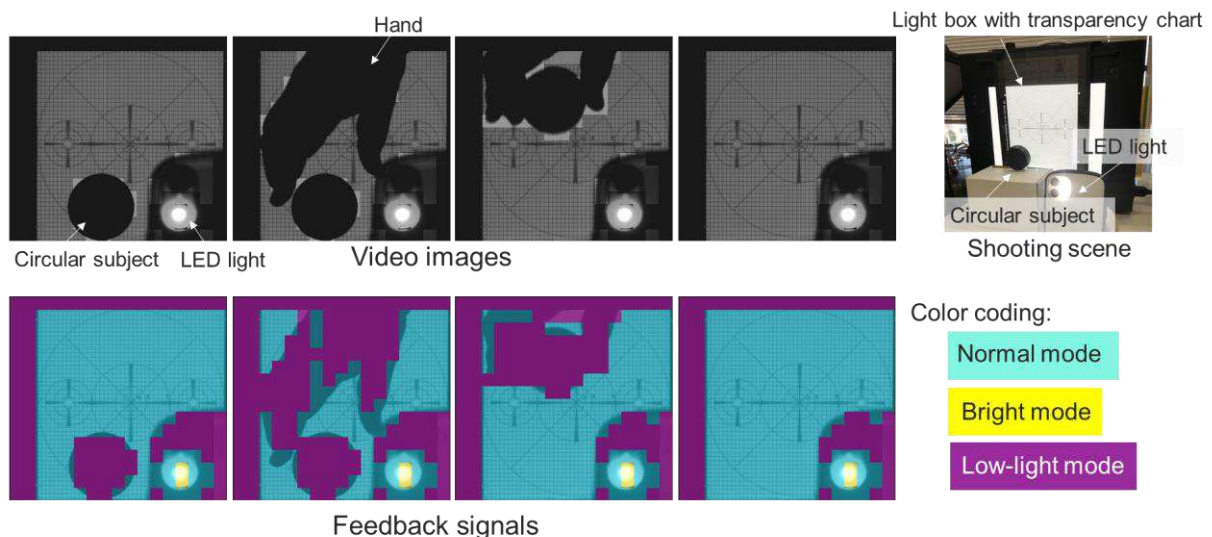
Fig. 7. Video images and feedback signals captured by the prototype scene-adaptive imaging system.

Analysis of Backside Illuminated CMOS pixels' Quantum Efficiency under Ultraviolet Illumination

N. Fassi[1,2], J.-P. Carrère[1], E. Leon Perez[1], M. Estribeau[2], P. Magnan[2] and V.Goiffon[2]

1. STMicroelectronics, 850 rue Jean Monnet, 38926, Crolles, France
2. ISAE-SUPAERO, Université de Toulouse, 10 av Edouard Belin, 31055, Toulouse, France
Email : {nour.fassi, jean-pierre.carrere, edgar.leonperez}@st.com { magali.estribeau, pierre.magnan, vincent.goiffon}@ isae-supaero.fr

Abstract—**The CMOS image sensor (CIS) market continues to grow and expand. One of the areas concerned by this expansion is UV applications. This work's main concern is to explain the performance of miniaturized backside-illuminated (BSI) CMOS pixels under UV light in the range of 200 nm to 400 nm. A previous study showed a quantum efficiency (QE) loss of nearly 50% between 400 nm (blue light) and 200 nm (UV-C).**

This study relies on optical and electrical simulations to understand the QE drop's causes for n-type photogate pixel. The optical transmission of the antireflective coating seems excellent under UV. Electrical simulations show that the electric field was high at the interface, allowing the photogenerated electrons to drift rapidly. Therefore, the hypothesis that links the QE loss to a recombination rate increase at the interface seems to fade, leading to new recombination mechanisms to investigate.

Keywords—**ultraviolet, backside illuminated pixels, CMOS image sensors, quantum efficiency, interface defects, antireflective coatings.**

I. Introduction

Ultraviolet (UV) detection [10 nm, 400 nm], is becoming more prominent in different areas, including forensics and environmental hazard detection [1]. However, to be able to detect UV, we are facing several challenges. First, according to Beer-Lambert law, 90% of UV light with a wavelength shorter than 300 nm is absorbed in the first 6 nm of Si (Fig. 1.)
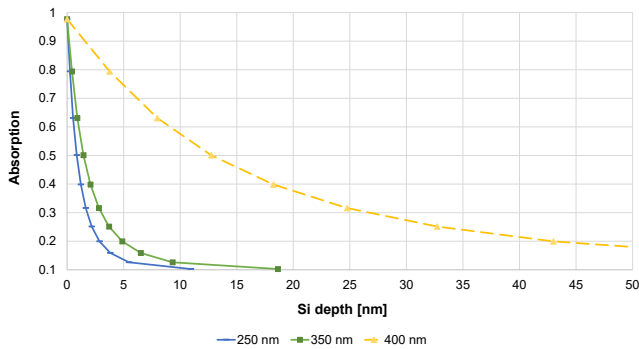
Fig. 1. UV light absorption in Si, according to Beer-lambert law

Another difficulty is the important role of Si/ SiO$_2$ interface defects [2] [3] on the charge collection efficiency.

The current work focuses on the absorption of Si and the influence of interface recombination on QE for UV wavelengths. As the wavelength becomes shorter, there may be a significant impact due to interface defects, which can increase with photon energy.

For this purpose, electro-optical simulations in visible light (VIS) and UV were carried out in the TCAD environment.

II. Experimental

The QE under UV of a p-type and a n-type BSI photogate (PG) pixel (Fig. 2.a&b) was measured.
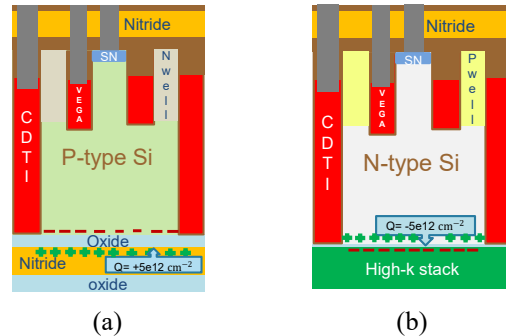
Fig. 2. Schematic cross section of a p-type PG (a) and an n-type PG (b), adapted from [4]

The reason for that choice is that the photogate has the possibility of being fully depleted. In theory, the PG can overcome this limited absorption depth of Si under UV without having to modify the Si epitaxy process to make it thinner.

The operating principle of the photogate pixel is simple: capacitive deep trench isolation (CDTI) offers fully depleted Si for carriers' storage, and dark current reduction. Also, PG uses epitaxial Si as active layer and its process doesn't involve any implantation, avoiding any interaction with process flow-related steps. On top of the device, front-side vertical Shallow Trench Transfer Gate and Planar Read Out transistor are implemented, and the well region resulting from the transistor realization works as top PG pinning layer. On the backside, the silicon surface passivation layer, also known as antireflective coating (ARC) is deposited, depending on the Si doping (n or p) [4].

Both investigated photogates give us the opportunity to have a closer look at two ARC stacks, namely high-k and oxide-nitride-oxide (ONO). Passivation of those pixels is ensured by inherent material fixed charges: for p-type pixels, a positively charged ONO stack, for which an electron-filled interface is needed, and a negative charged high-k stack (with high dielectric constant materials), for n-type pixels [5]

It is worth stating that measurements were carried out on pixels without micro-lenses.

Next, quantum efficiency measurements were carried out using a monochromator and a xenon lamp to measure QE down to 200 nm with a sampling step of 10 nm, at 40°C, with a f number of f/2 and a slit aperture of 15 nm [6].

Electro-optical simulations are performed using 3D TCAD tools from Synopsis.

III. RESULTS

A. Optical measurements

Although the photogate pixel is optimized for visible light, it shows a quite good QE response in the UV, comparable to the literature. However, both types of photogates show an approximately 50% decrease while the wavelength varies from 400 nm to 300 nm, as shown in (Fig. 3.)

To understand this trend, investigations should focus on whether this is due to phenomena related to the silicon or silicon oxide interface or to the anti-reflective coatings (ARC) absorption, the ARC being a back interface passivation, preventing recombination, depends on the nature of the silicon interface.
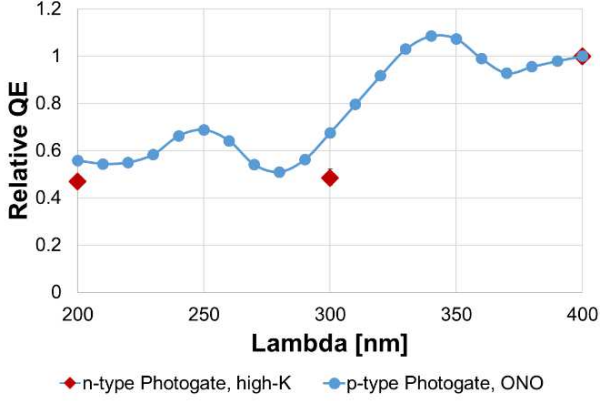
Fig. 3. Relative QE of n-type and p-type photogates under UV normalized at 400 nm

IV. SIMULATIONS

A. Optical simulations

Fig. 4 shows optical transmission simulation of the two ARC stacks used in the two PG pixels: the simulated transmission remains high under UV and does not explain the QE loss going down to 200nm.
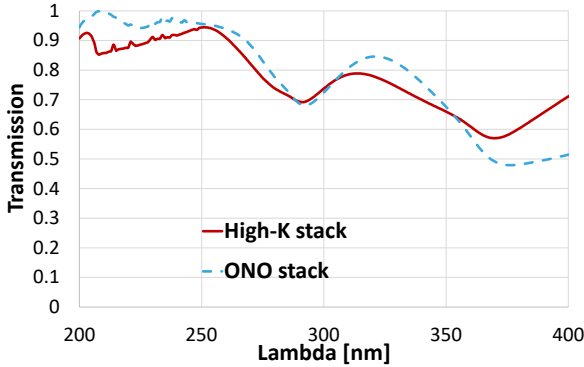
Fig. 4. Simulation of pixels' ARC transmittance, UV wavelengths

B. Electrical simulations

The electrostatic comparison of n-type and p-type PG is presented in (Fig. 5), where the potential map looks rather similar.
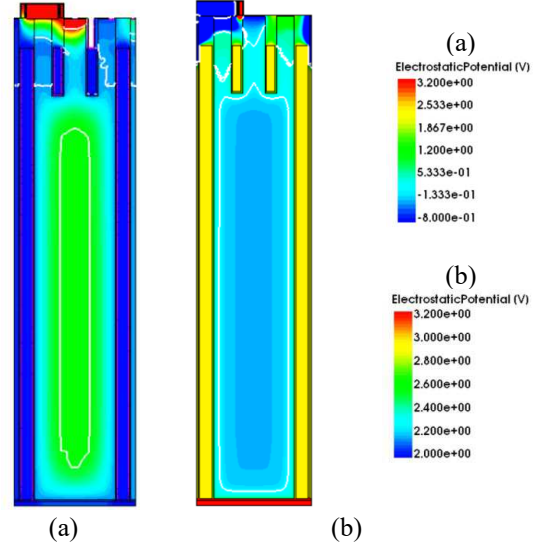
Fig. 5. Cross section showing electrostatic potential of n-type (a) and p-type (b) photogates

For the next simulation, focus is on the n-type PG because its ARC stack's inherent charge induces a higher electric field (as visible in Fig. 6). We can also notice that for the same fixed charge, the interface's electric field is much higher for n-type pixels than the p-type ones.
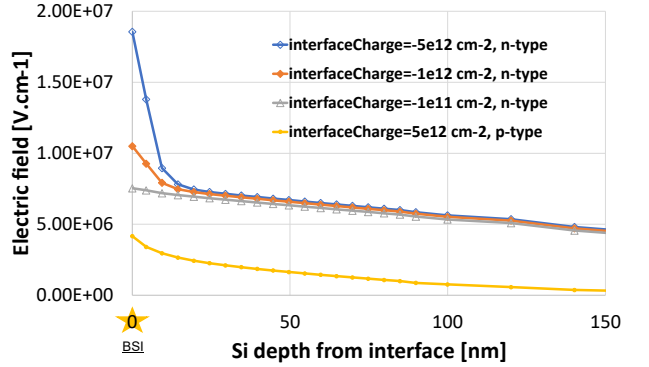
Fig. 6. The first 0.15 μm of Si from the backside, to see the impact of different interfaces or interface's charge on electric field.

Indeed, the interface charge does not influence much the deep potential profile (see Fig. 7) but brings out a little more electric field in the backside (as shown in Fig. 6).

The next electrical behavior simulation is of an n-type PG pixel (Fig. 2.b), to study the behavior of the photogenerated charges close to the backside interface.
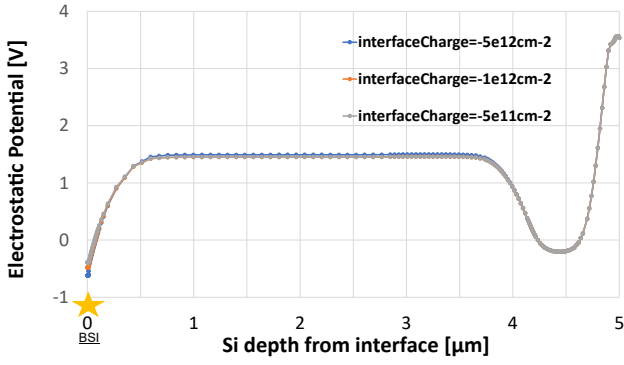
Fig. 7. Impact of the backside interface charges on the electrostatic potential of a n-type photogate pixel.

C. Illumination simulation

Illuminations under different wavelengths have been simulated using the 3D TCAD tools and its optical generation function, as (Fig. 8) illustrates. For each wavelength, the number of incident photons has been kept constant, corresponding to the experimental conditions of the QE measurement. The quantum yield (QY) of the pixel has next been calculated as the ratio of collected electrons over the number of optically photogenerated electrons inside the pixel silicon volume.

(a)   (b)
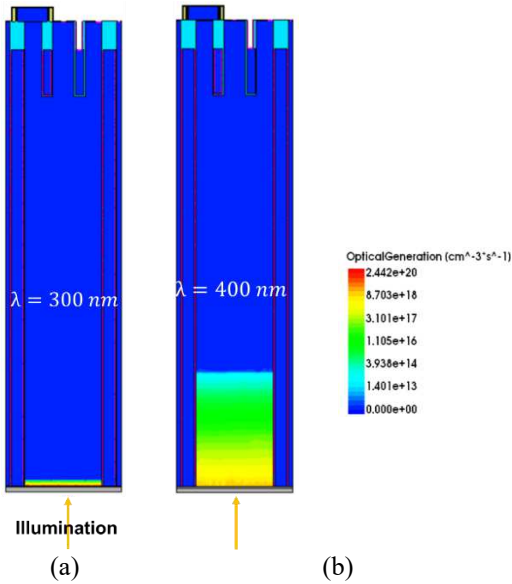
Fig. 8. Cross section of optical generation simulation in a n-type PG, illuminated from backside at (a) $\lambda = 300 \ nm$ and (b) $\lambda = 400 \ nm$

The simulations allow for solving Poisson and continuity equations for drift-diffusion of carriers and Shockley-Read-Hall (SRH) recombination using the Scharfetter model. Trap density levels within silicon were introduced in a second step.

The Table I summarizes the QY simulation checks without trap: we verify that the QY is about 100%, and hence all the photogenerated charges are collected.

Fig. 9 shows the photogenerated electrons and their distribution among the silicon depths, for 200nm, 300 nm and 400 nm UV illuminations. Total generated charges were normalized to 100 electrons for the whole pixel.
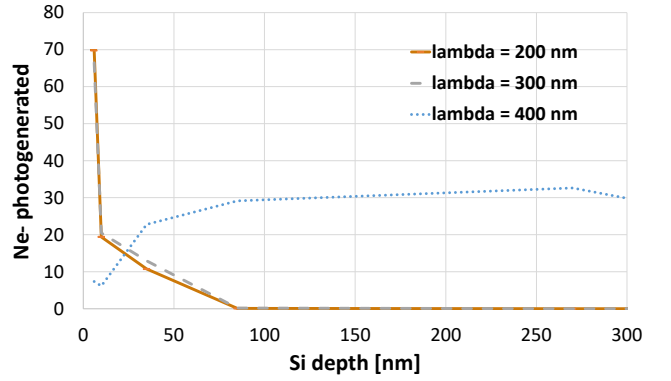
Fig. 9. Photogenerated electrons in relative values, depending on the Si depth in the pixel and the illumination wavelength, no traps added to the simulation.

We can observe that, indeed, most electrons are generated in the first few nanometers of bulk silicon, especially for wavelengths shorter than 300 nm. For 400nm wavelength illumination, the electrons are generated deeper, between 100 nm and 300 nm from the interface.

Once collected by the photogate electric field, the photogenerated electrons are located in the well of the center of the structure, as shown by the electrostatic potential in Fig.7.

D. Illumination simulation with interface traps

Next, traps have been introduced on the backside interface to study their impact on the photogenerated charges. The trap's characteristics are donor in our case, and the cross section is about 5.10^{-13} cm^{-2} in this simulation; the traps density has been modulated.

TABLE I SIMULATED QUANTUM YIELD OF THE PIXEL, WITH AND WITHOUT BACKSIDE INTERFACE TRAPS

λ [nm]	Trap type	QY
400	No traps	100%
	Traps, donor Dit = 2e11 cm^{-2}	~100%
	Traps, donor Dit = 2e12 cm^{-2}	~100 %
300	No traps	100%
	Traps, donor Dit = 2e11 cm^{-2}	~100%
	Traps, donor Dit = 2e12 cm^{-2}	~100%
200	No traps	100%
	Traps, donor Dit = 2e11 cm^{-2}	~100%
	Traps, donor Dit = 2e12 cm^{-2}	~100%

We observe in Table 1 that the introduction of the backside interface traps has no significant impact on the pixel's internal Quantum efficiency. This suggests that the recombination effect of the photogenerated electrons on the backside interface traps does not explain the QE decrease from 400 to 200nm.

V. DISCUSSION

As observed from the measurements and optical simulations, ARC stacks can indeed modulate the QE appearance, but not below 300nm, because they show good transmission under UV.

What follows is a plausible explanation for the excellent electrical quantum yield of the pixel. We must take into consideration that the interface electric field can help electrons drift quicker to the depletion zone or even to the collection zone. Fig. 6 shows that the ARC fixed charges account for a

strong interface electric field, which can reach 10^7 V.cm^{-1}. With such electric field, the photogenerated electrons drift velocity µ.E can reach high values close to the silicon saturation v_s about 10^7 cm/s.

With this drift velocity, the time spent at the interface can now be estimated by the electrons generated in the first ten nanometers, which will be less than 1ps, not letting much trapping impact probability to happen. Especially if we consider that the surface recombination velocity is between 200 and 300 cm.s^{-1} [7] which leads to a recombination time way bigger than the speed at which the interface is crossed.

Finally, the backside drift velocity in the pixel seems to be optimum to collect UV photogenerated carriers at the interface, preventing the sensitivity loss due to carriers' recombination at interface, as illustrated in Fig. 10.
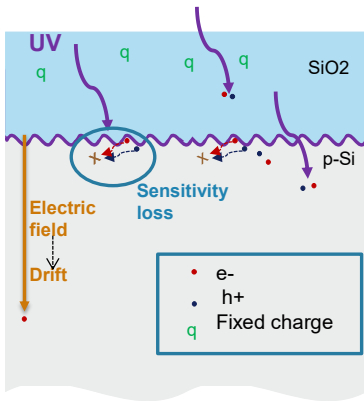
Fig. 10. Schematic illustration of a pixel's cross section with UV-induced oxide charging effects, inspired from [2]

Many plausible explanations for QE loss are discussed below. First, the other pixel interface may have a role in the carriers recombination effect, like the sidewall interface of the CDTI trench. Indeed, carriers generated very close to the backside interface may show more sensitivity during their transport to SRH recombination occurring at the sidewall interface traps.

Next, other recombination mechanisms may occur with such a high carrier density, like band-to-band recombination, where energy is exchanged in a radiative or Auger process.

This very high density of carriers at the interface is most likely a factor that increases the capture probability and so decreases the recombination lifetime, as demonstrated by (1), A, B & C being defined in [8] n_d is the defect density and $n_{s,s',\alpha}$ is electrons density.

$$R_{ec} \approx \frac{2\pi}{\hbar} n_d \times A \times \left[n_{s,s',\alpha} + B\right] \times C \qquad (1)$$

Hence a good part of the photogenerated electrons probably drifts away thanks to the strong electric field at the interface, but part of them remains and suffers the effects of recombination at the interface, which degrades the QE.

These stated effects were not considered in the presented simulations and will be investigated in further work. The optical behavior of the antireflective stack will also be compared with 3D Lumerical simulations.

VI. CONCLUSION

Quantum efficiency (QE) on different BSI pixels have been measured under UV. Despite a quite good response in UV, a loss at shorter wavelengths was noticed, and the inducing phenomena were discussed.

ARC stacks transmission seems to modulate the QE, but not under 300 nm. So, it should not be a big contributor, as it has good transmission in UV.

Backside traps SRH recombination does not seem either to be the first cause of this QE loss, due to the high electric field at this interface. Finally, the main causes of the QE loss at short wavelength are most probably a recombination enhanced by the very high density of photogenerated carriers at the interface, or recombination during the carriers transport.

REFERENCES

[1] Okino, T., Yamahira, S., Yamada, S., Hirose, Y., Odagawa, A., Kato, Y., & Tanaka, T. (2018). A real-time ultraviolet radiation imaging system using an organic photoconductive image sensor. Sensors, 18(1), 314.

[2] Nakazawa, T., Kuroda, R., Koda, Y., & Sugawa, S. (2012, February). Photodiode dopant structure with atomically flat Si surface for high-sensitivity and stability to UV light. In Sensors, Cameras, and Systems for Industrial and Scientific Applications XIII (Vol. 8298, pp. 186-193). SPIE.

[3] LI, Flora et NATHAN, Arokia. CCD image sensors in deep-ultraviolet: degradation behavior and damage mechanisms. Springer Science & Business Media, 2005.

[4] Roy F, Suler A, Dalleau T, Duru R, Benoit D, Arnaud J, Cazaux Y, Chaton C, Montes L, Morfouli P, Lu G-N. Fully Depleted, Trench-Pinned Photo Gate for CMOS Image Sensor Applications. Sensors. 2020; 20(3):727. https://doi.org/10.3390/s20030727

[5] SACCHETTINI, Y., CARRÈRE, J.-P., DOYEN, C., et al. A highly reliable back side illuminated pixel against plasma induced damage. In : 2019 IEEE International Electron Devices Meeting (IEDM). IEEE, 2019. p. 16.5. 1-16.5. 4.

[6] Fassi, N, Carrère, JP., Estribeau, M., Goiffon, Vincent. (2023, January). Quantum efficiency of various miniaturized backside illuminated CMOS pixels under ultraviolet illumination. In Proc. IS&T Electronic Imaging.

[7] Eades, W. D., & Swanson, R. M. (1985). Calculation of surface generation and recombination velocities at the Si-SiO2 interface. Journal of applied Physics, 58(11), 4267-4276.s, W. D., & Swanson, R. M. (1985). Journal of applied Physics, 58(11), 4267-4276.

[8] Wang, H., Strait, J. H., Zhang, C., Chan, W., Manolatou, C., Tiwari, S., & Rana, F. (2015). Fast exciton annihilation by capture of electrons or holes by defects via Auger scattering in monolayer metal dichalcogenides. Physical Review B, 91(16), 165411.

Near Infrared Quantum Efficiency Simulations for CMOS Image Sensors

Andrew Perkins
Intelligent Sensing Group
onsemi
Nampa, ID USA
andrew.perkins@onsemi.com

Swarnal Borthakur
Intelligent Sensing Group
onsemi
Nampa, ID USA
swarnal.borthakur@onsemi.com

Abstract— **This paper presents results of 3D finite-difference-time-domain (FDTD) simulations for predicting Near Infrared (NIR) Quantum Efficiency (QE) of complementary metal oxide semiconductor (CMOS) image sensors. JMP statistics software [1] and Lumerical FDTD [2] simulation software were used to setup and analyse a Central Composite Design (CCD) design of experiment (DOE) with factors of pixel pitch varying from 1.5um to 3.5um, epitaxial silicon substrate thickness varying from 3um to 8um, and number of inverted pyramids per pixel varying from 1 to 36. NIR QE at wavelengths of 850nm and 940nm is predicted and compared to products on the market. Factors to further increase NIR QE are also discussed.**

Keywords—near infrared, quantum efficiency, inverted pyramids, CMOS image sensor

I. INTRODUCTION

Near Infrared (NIR) quantum efficiency (QE), specifically at wavelengths of 850nm and 940nm, is becoming increasingly important in CMOS Image Sensors (CIS). The need for enhanced NIR QE is driven by multiple markets: the security market wants to reduce the power consumption of active light emitting diodes (LED) illuminants; the automotive market wishes to employ NIR in-cabin monitoring for driver awareness, and aid in advanced driver assistance systems (ADAS) detection of objects at night; the industrial market wants NIR for automated inspection with enhanced contrast under low-light conditions and improved spectral information; and the consumer markets need NIR for facial recognition, depth detection, and internet of things (IoT). The current NIR QE of standard CMOS image sensors (~3um epitaxial silicon thickness) without any NIR enhancement structures is 18%/9% at 850/940nm respectively, but NIR QE of up to 60%/40% at 850/940nm is now needed to stay competitive, with QE of 70%/50% at 850/940nm desirable.

To keep manufacturing costs low, silicon remains the substrate of choice, compared to Germanium or III-V substrates. Silicon substrate enables the use of existing CMOS manufacturing techniques. However, silicon is an indirect bandgap semiconductor and has a weak absorption coefficient of <0.1 um^{-1} in wavelengths >800nm. This means the Si thickness needs to be at least 23um thick to achieve 70% QE at 850nm wavelength, and 38um thick to achieve 50% QE at 940nm wavelength as shown in Fig.1. Current costs and manufacturing make such a thickness impractical, especially for backside illumination (BSI) image sensors. This means a solution beyond simply increasing the silicon substrate thickness must be utilized. NIR light scattering structures are a way to effectively increase the absorption path length.
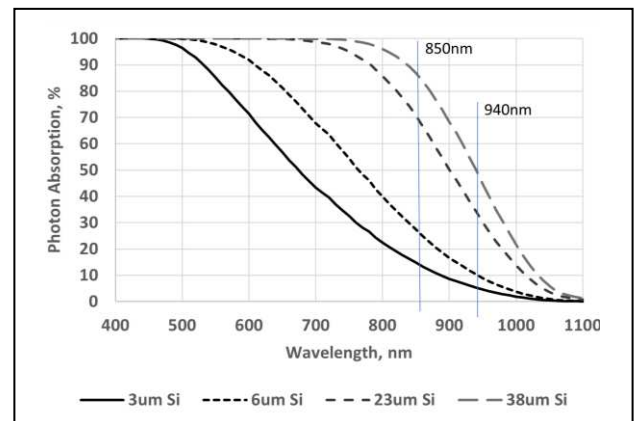

Fig. 1. Photon Absorption at silicon thicknesses of 3um, 6um, 23um, and 38um (based on absorption coefficient data [3])

II. CMOS IMAGE SENSOR WITH NIR LIGHT SCATTERING STRUCTURES

BSI has enabled integration of light scattering surfaces directly on the silicon surface where light enters the pixel. CIS manufacturers have increased NIR QE by incorporating light scattering structures like refractive inverted pyramids (IVP) or diffractive trenches on the silicon surface.

A. Inverted Pyramids (IVPs)

IVPs are formed using a wet etch process that follows the (111) crystalline plane of silicon. This results in fixed relationship of pyramid width and height, and an etch angle of ~54.7° when etched in a (100) silicon surface. Because the wet etch slows on the (111) silicon crystal plane, the orientation of the wafer notch relative to the crystal orientation is important to consider. Depending upon the wafer notch orientation relative to crystal orientation, an IVP will etch with walls parallel to the pixel (square) if the notch is at <110> direction, or with pyramid walls at an angle to the pixel (diagonal) if the notch is at <100> direction. This can impact the pyramids layout and scattering effectiveness.

B. Diffractive Trenches

Diffractive trenches are formed with a dry etch process during backside processing and are typically ~100-200nm wide with a depth of 400-1000nm. A dry etch requires an annealing to reduce the damage to silicon surface during the dry etch. The geometry is largely guided by subwavelength

diffractive physics. The geometry and layout of trenches are not limited by the crystal orientation like IVPs.

C. Other relevant structures

Other structures such as microlens, color filter array (CFA), anti-reflective coating (ARC) films, backside or frontside trench isolation (BDTI/FDTI), and metal structures and circuitry can impact NIR QE as shown in Fig. 2. These other structures can lead to further optimization of NIR QE.

Table 1 shows a sample of NIR enhanced CIS with a variety of pixel pitches, silicon thicknesses, and NIR scattering structures. When considering the design of an image sensor, it would be helpful to be able to estimate what NIR QE to expect. Comprehensive three-dimensional Finite-Difference-Time-Domain (FDTD) simulations along with design are used to model NIR QE response as a function of pixel pitch, silicon thickness, and IVP layout.

III. SIMULATIONS

3D Finite-Difference-Time-Domain (FDTD) optical simulations using Lumerical [2] software with relevant geometry and material properties is used to simulate NIR QE.

The importance of how NIR scattering structures play a role in NIR QE can be seen in Fig. 3 which shows the QE results of a 3D FDTD simulations for an image sensor with 2um pixel pitch, 6um silicon thickness, and 4um backside deep trench isolation (BDTI). The simulations include results for no NIR scattering structures, IVP in a diamond configuration, IVP in a square configuration, and trenches in a 'star' configuration.
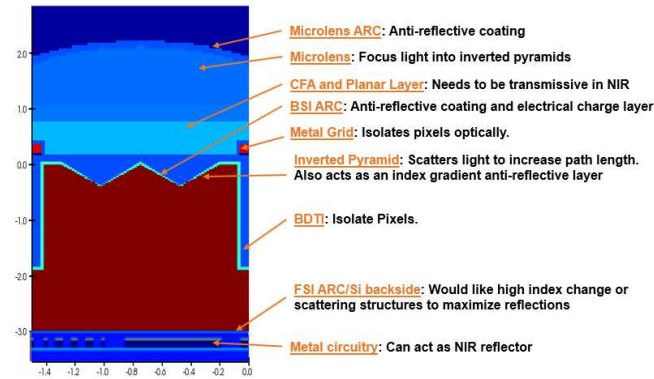
Fig. 2. 3D FDTD simulation cross-section with important structures for NIR QE
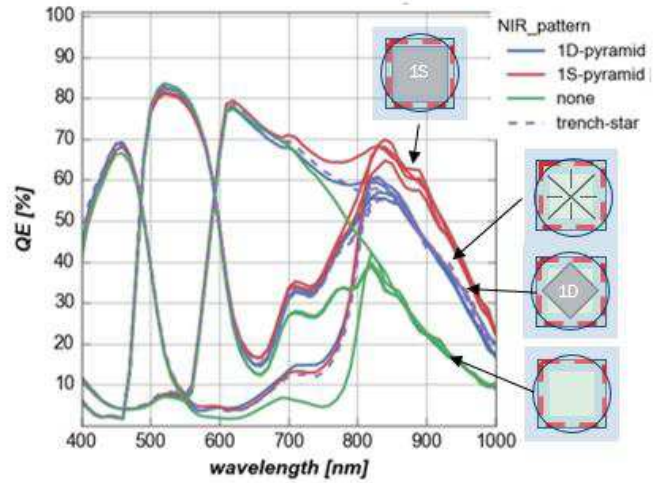
NIR type	NIR width [nm]	QE%				
		850nm	940nm	Blue	Green	Red
1S: Square IVP	1414	66	44	68	82	79
1D: Diagonal IVP	910	56	35	69	82	77
trench-star	120	57	37	69	83	78
None	0	33	16	66	84	78

Fig. 3. Simulated QE for 2um pixel pitch, 6um Si thickness, 4um BDTI depth with various light scattering structures

IVPs and trenches can have similar NIR QE performance. It can be seen from Fig.3 that an IVP oriented in a diamond configuration (1D), and trenches in a star pattern have similar NIR QE. However, if the pyramid is etched in a square configuration (1S) then NIR QE increases due to more light entering the pyramid and being scattered more effectively. This study will focus on IVPs in a square orientation.

IV. DOE AND RESULTS

JMP software is used to develop a modified Central Composite Design (CCD) design of experiment (DOE) with factors of pixel pitch (1.5um to 3.5um), NIR layout (1 to 6 IVPs per pixel side), and silicon thickness (3 to 8um). Fig. 4 shows the DOE values with the simulated NIR QE results, and illustrations of how IVPs are laid out.

TABLE I. SAMPLE OF IMAGE SENSORS WITH NIR SURFACE SCATTERING STRUCTURES

|	Sony IMX332[4]	SmartSens SC5035 [4]	OmniVision 0S05A20 [5]	OmniVision OS02C10 [4]	Samsung (Conference) [6]	Sony IMX462 [4]	onsemi AR0830 [7]
Date Announced/Published	2017	2017	2018	2020	2020	2021	2022
Pixel Pitch (um)	1.12	2.0	2.0	2.9	2.3	2.9	1.4
Si Thickness (um)	3.5	2.4	3.6	6	8	6	6
BDTI depth (um)	2 (backside)	1.6 (backside)	2.1 (backside)	4 (backside)	8 (frontside)	2 (backside)	4 (backside)
NIR type	IVP	IVP	Trenches	IVP	Trench-Star	IVP	IVP
NIR Width (nm)	400	770	160	580	170	400	900
NIR Depth (nm)	236	545	410	360	1000	236	636
# NIR Structures per pixel	4	4	9	16	1	36	1
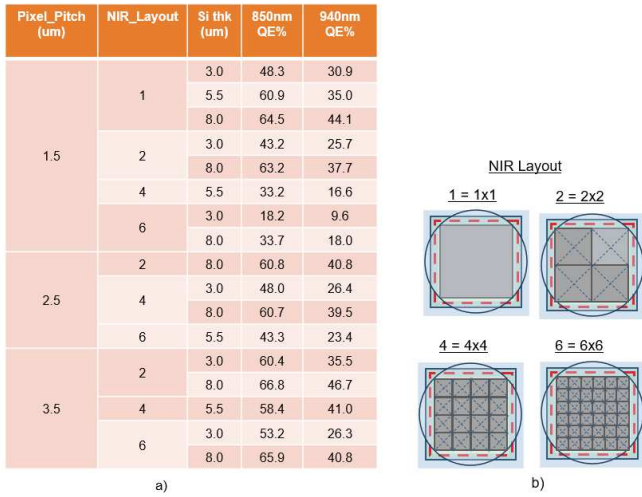
Fig. 4. a) Modified Central Composite Design of Experiment with results of NIR QE. b) NIR pyramid layout

A least squares fit analysis was run in JMP to determine the important factors. Fig. 5a shows the important factors (p-value <0.05) are pixel pitch, silicon thickness, NIR_layout, and the interaction term of pixel pitch and NIR_layout. The prediction profiler and interaction profiles as shown in Fig. 5b,c can be used to estimate NIR QE with pixel pitch, number of NIR pyramids, and silicon thickness as inputs.

In general, the following observations can be made:

1) NIR QE increases linearly with pixel pitch. This is to be expected as absorption path increases due to lateral scattering of light.
2) Single large pyramids are preferred over a greater number of smaller pyramids. As the width of pyramid increases it becomes easier to focus into the pyramid and more light is effectively scattered. Pixel pitch plays a role in the practical size and layout of IVPs.
3) As expected, NIR QE increases with silicon thickness. For example, given a 2um pixel pitch and four IVPs, the expected QE@940nm for 3um Si thickness would be 29%, and for 6um Si thickness QE@940nm would be 36%. For a target of 40% QE@940nm then 7.4um of Si thickness would be needed.

In this study the microlens size and shape was a fixed function of pixel pitch. Further NIR QE improvement is possible by optimizing the microlens for each specific pixel pitch and NIR layout.

V. CONCLUSIONS

3D FDTD simulations and DOE can provide guidance and insight when estimating NIR QE for an image sensor for a given pixel pitch, NIR layout, and silicon thickness.

ACKNOWLEDGMENT
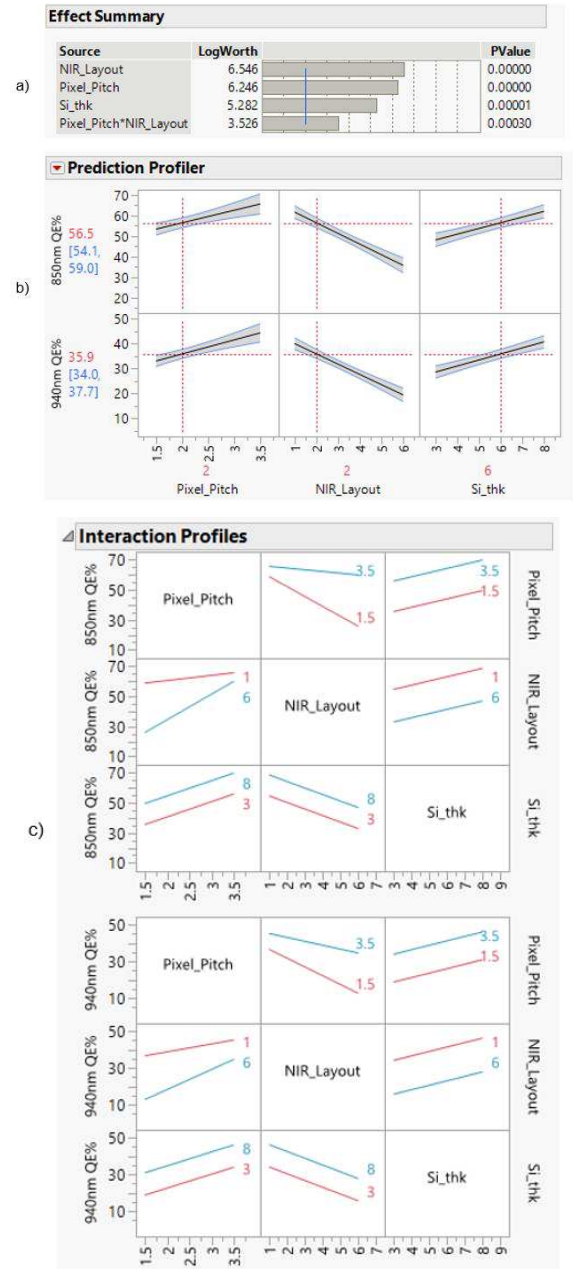
Fig. 5. A) Least squares fit. b) prediction profiler c) interaction profiles for NIR QE based on pixel pitch, NIR layout, and silicon thickness

REFERENCES

[1] JMP 14.3. SAS Institute Inc., Cary, NC, 1989-2021.

[2] ANSYS Lumerical FDTD, Release 2022R1.4

[3] Martin A. Green, 'Self-consistent optical parameters of intrinsic silicon at 300 K including temperature coefficients', *Solar Energy Materials and Solar Cells* **92** (11), pp.1305-1310, (2008)

[4] Tech Insights, 2021 Image Sensor Device Essentials (DEF) Annual Seminar. 2021.(BRF-2201-808)

[5] TechInsights, OmniVision OS05A20 Device Essentials Plus Summary. DEP-1810-801. 2018.

[6] J. -K. Lee *et al.*, "5.5 A 2.1e⁻ Temporal Noise and −105dB Parasitic Light Sensitivity Backside-Illuminated 2.3µm-Pixel Voltage-Domain Global Shutter CMOS Image Sensor Using High-Capacity DRAM Capacitor Technology," *2020 IEEE International Solid- State Circuits Conference - (ISSCC)*, 2020, pp. 102-104, doi: 10.1109/ISSCC19947.2020.9063092.

[7] O. Skorka, S. Micinski, A. Perkins, B. Gravelle, X. Li, R. Ispasoiu, "1.4um pixel, 8MPixel, thick epi image sensor for RGB-IR imaging," 2021 International Image Sensor Workshop, 2021, R32.

Metasurface-based planar microlenses integrated on back-side illuminated CMOS pixels

1st Martin Lepers
STMicroelectronics
CRHEA
Grenoble, France
martin.lepers@st.com

2nd Alain Ostrovsky
STMicroelectronics
Grenoble, France
alain.ostrovsky@st.com

3rd Patrice Genevet
Univ. de Côte d'Azur
CRHEA
Valbonne, France
patrice.genevet@crhea.cnrs.fr

4th Stéphane Lanteri
Univ. de Côte d'Azur
Inria, CNRS, LJAD
Sophia Antipolis, France
stephane.lanteri@inria.fr

5th Jérôme Vaillant
Univ. Grenoble Alpes
CEA-Leti
Grenoble, France
jerome.vaillant@cea.fr

Abstract—*Metasurface based microlenses (or metalenses) are planar microlenses which could be an alternative to replace refractive microlenses made of resist on image sensor. This paper aims at exploring through simulation the integration of metalenses on CMOS pixel array. Theses simulations precede metalenses implementation on STMicroelectronics 2.8, 4 and 8 µm pixels with updated fabrication process involving immersion photolithography. Emphasis is given on improving metalenses performances with consideration to fabrication capacities available.*

Index Terms—*CMOS, BSI, metasurface, microlens, metalens, immersion photolithography*

I. Introduction

Based on prior work on FSI planar microlenses to improve the sensitivity of Single Photon Avalanche Diode SPAD [1], [2] further studies are being conducted to adapt the technology on CMOS back-side illuminated pixel[1].

Metasurfaces are nanostructured surfaces capable of manipulating light at a sub-wavelength scale to achieve a macroscopic optical function [3]–[6]. The microlenses designs discussed in this article are based on metasurfaces encoded with a phase profile allowing the concentration of light at a given focal length. They are composed of sub-wavelength cylindrical pillars of high-index material embedded in a low-index material substrate. By modulating the pillars diameter it is possible to induce localized phase-shift which could then be spatially distributed to replicate the phase profile of a lens.

II. Fabrication Process

Demonstrators on industrial CMOS image sensors are currently under fabrication and characterization will be available in the second half of 2023. This generation of metalenses is fabricated with an updated version of the previous generation process [2]: 193 nm immersion photolithography and improved OPC (Optical Proximity Correction) are used to achieve more aggressive design (smaller minimum pillar diameter) and better control of pillar size. As of today, the fabrication process started and metrology tools such as SEM are used to control the fabrication (see Fig. 1).

Fig. 1: Tilted SEM view of a 2×2 metalenses array made with immersion photolithography tool.

III. Simulation

In the following section we examine different design approaches for metalenses. At first we will discuss the metalenses geometry with consideration to their focalization properties and then present recent work about phase profile engineering.

A. Simulation workflow

The metalenses discussed in this paper follow the design method described in [1], [2]. The phase profile of a perfect lens is sampled with unitary pillars (meta-atoms or nanopillars) of various diameters inducing local phase-shift. Those meta-atoms are arranged in 2D space to reproduce the targeted phase profile. The set of parameters (pitch, height, radius) defining the structure of the pillars is referred as "library" (see Fig. 2) and is obtained through FDTD[2] (Lumerical FDTD) and RCWA[3] (Reticolo) simulations.

To evaluate the performance of a design, the optical stack is simplified to its essentials features and simulated with

---

[1]The design wavelength for the microlenses based on metasurface is 940 nm

[2]Finite Difference Time Domain
[3]Rigorous Couple Wave Analysis

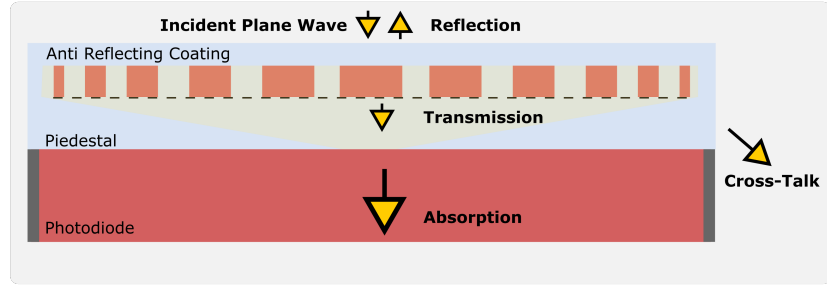## Library Parameters
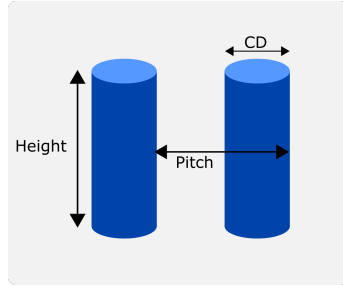
Metalens Parameters

Fig. 2: Schematic of library parameters (left) and description of a metalens integrated on a simplified pixel (right).
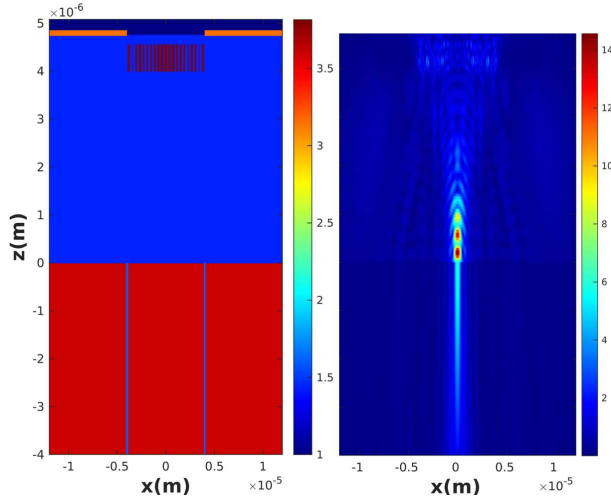

Fig. 3: (Left) Refractive index visualization of the simulation (Right) Electric Field (V/m) in the simulation. We simulate a square metalens of 8 μm width with a focal length of 4 μm on top of a 3×3 pixels (red) array isolated from each other by isolation trench (vertical blue lines). The lens is embedded in silicon oxyde (blue) and tungsten shielding (orange) is added to estimate cross-talk.

Lumerical-FDTD (see Fig. 3). The studied metalens is embedded in silicon dioxide and lays on a pedestal which height is a design parameter. Another layer of oxide is added on top of the stack to reduce reflection. The photodiode is surrounded by isolation trenches which goal is to reduce cross-talk. In the simulation a shielding[4] (tungsten) is added to evaluate the light loss due to diffraction outside the targeted pixel which is also referred as optical cross-talk. For each simulation the aperture size in the shielding is equal to the width of the metalens.

B. Microlens geometry

Simulations are performed on metalenses that have the same size as the pixels they are placed on, which is the most commonly used configuration for imaging devices. Focus is given on the amount of light transmitted to the photodiode as we consider total absorption in the silicon and zero cross-talk due to the isolation trenches.

We calculate the metalenses ensquared energy (see Fig. 4) in the focal plane to assess their focusing efficiency. The

[4] It is important to note that shielding will not be present on the fabricated structure but the aperture does create additional diffraction in the simulation
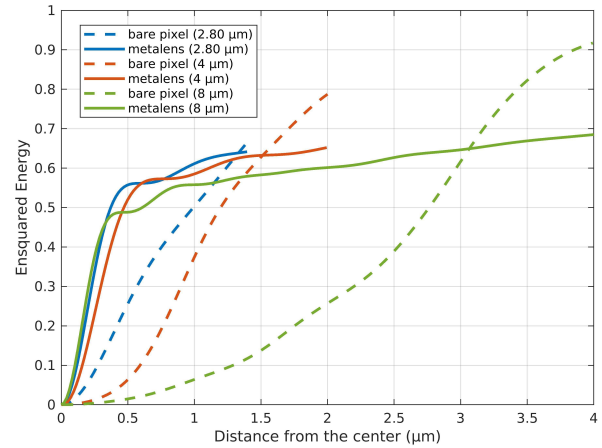
Fig. 4: Ensquared energy comparison between a pixel with metalens (solid line) and a pixel without metalens (dashed line).

calculation involves integrating the Poynting vector flux over an increasing square surface area. We observe an intensity distribution in the focal plane following the profil of an altered Airy disk. We consider the focusing efficiency as the ratio between intensity inside a square of width equals to the diameter of the Airy spot first dark ring (which we estimated to be 0.5 μm) and the intensity collected by the total surface area. For pixel size of 2.80, 4 and 8 μm with a fixed focal length of 4 μm we find the associated focusing efficiency of $87\,\%$, 79% and 71%. As metasurface allows arbitrary manipulation of phase through the variation of pillar diameter, metalenses enables focalization properties out of reach for melted refractive microlenses limited by the fabrication process. Low numerical-aperture and off-axis [2] metalenses have been simulated on pixels array of 2×2, 3×3 and 4×4 pixels to assess their performances (see Fig. 5 to Fig. 7).

Improvement on the library [2] (denoted library 1, pitch=370 nm, height=350 nm) used to design metalenses on FSI SPAD was also studied through multiple simulations at the nanopillar and at the metalens scale. A more compact library (denoted library 2, pitch=300 nm, height=750 nm) demonstrated better performances than the previously used set of pillars. This library have been selected after running simulations of nanopillars through most of the parameter space considering the updated fabrication process. It has been
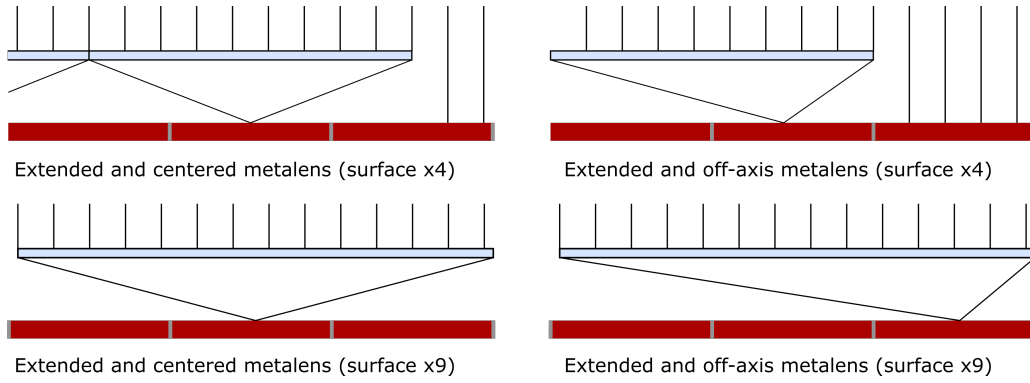
Fig. 5: Schematic of an extanded metalens and an off-axis metalens positioned over an array of 2×2 and 3×3 pixels.

selected for its transmission and focalization performances given the considered size and focal length of the metalens. As the surface of the lens expands, we observe continuous increase of the intensity transmitted to the photodiode with a maximum gain of 9.5 for a metalens with a surface area 16 times larger than the pixel. For off-axis metalenses the simulations show a maximal gain of 6.85 for the same surface extension. For each design variation, the library 2 increases the transmission to the targeted photodiode in comparison to library 1. A maximum gain of 1.31 is achieved between the two centered metalenses with a surface area 16 times the area of the photodiode.
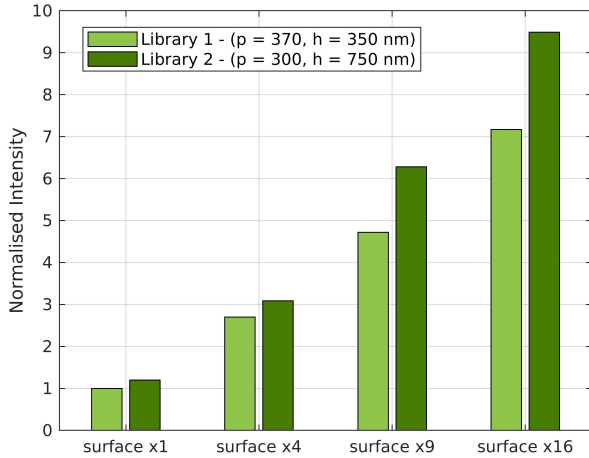
Fig. 6: Centered metalens simulated transmission for various footprint surface. The surface of the metalenses increases from left to right and both library 1 (light green) and library 2 (dark green) are used. Focal length is 4 μm.

Simulations show an overall performances increase with the compact library with lower reflection and cross-talk than the one used previously. It may be explained by a better sampling of the phase profile allowed by the smaller pitch.

C. Phase profile investigation

Most metalenses designs in literature are based on a hyperbolic phase profile (1), in this section, we evaluate the use of a quadratic phase profile (2). Quadratic phase profile translates a linear phase (which corresponds to a plane wave with oblique incidence) into spatial shift and so the effect of a
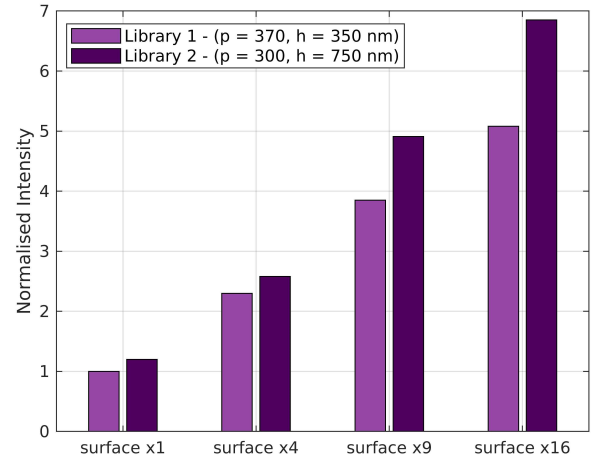
Fig. 7: Off-Axis metalens simulated transmission for various footprint surface. The surface of the metalenses increases from left to right and both library 1 (light purple) and library 2 (dark purple) are used. Focal length is 4 μm.

non-zero incidence would be a spatial translation of the focal spot whereas it would spread with a hyperbolic phase profile.

$$\phi_{hyp}(r) = -k_0 n(\sqrt{f^2 + r^2} - f) \tag{1}$$

$$\phi_{quad} = \frac{-k_0 n}{2f} r^2 \tag{2}$$

with k_0 the wavevector, n the index of the substrate, f the encoded focal length and r the radial coordinate on metasurface plane. Following the work on wide FOV metalenses [7], we implement on the pixel stack previously described a hyperbolic metalens and a quadratic metalens with a focal length f = 4 μm and a size of D = 4 μm. As the incident angle varies it is expected to observe a lower expansion of the focal spot for the quadratic lens in comparison of the hyperbolic lens and therefore to achieve a better angular tolerance.

Simulations show that the FWHM (Full Width at Half Maximum) of both metalenses focal spots increases with incident angle (see Fig. 8). However as the incident angle increases, the quadratic phase profile shows a better resilience to maintain the expansion of the focal spot with a 40 nm difference for a angle of $45°$. As expected the quadratic metalens presents a bigger focal spot at normal incidence
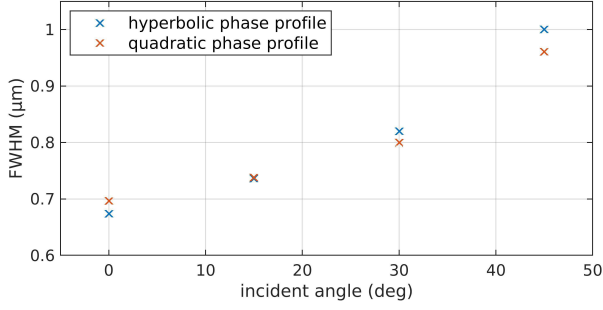
Fig. 8: FWHM variation of the focal spot with incident angle for a lens designed with a hyperbolic phase profile (blue) and with a quadratic phase profile (orange).

Fig. 9: Performance comparison between a hyperbolic (blue) and an effective-index based metalens (orange).

as quadratic phase profile introduces spherical aberrations. These simulations results suggest that the phase profile may be optimized at the metalens scale to design angle resilient metalenses for imaging system.

D. Phase profile encoding

Different approachs can be used to design phase optics based on metasurfaces. Previously we referred to the look-up table method to create libraries which consist in running simulation of unitary pillar, extract the induced phase-shift, and use them as a database to sample the targeted phase profile. Research on "CMOS-compatible all-dielectric metalens" [8] presents a different design approach based on an effective-index n_{eff} modulation and Maxwell-Garnett mixing rule. We propose here a performance comparison between those designs of metalenses. Pillar pitch and height is fixed at p=370 nm and h=550 nm but for the effective-index metalens, pillar diameter D is calculated analytically with the equation (3) and (4).

$$D(r) = \sqrt{\frac{4}{\pi}p^2 F(r)} \quad (3)$$

$$F(r) = \frac{n_{eff}^2(r) - n_0^2}{n_{eff}^2(r) + n_0^2}\frac{n_p^2 + n_0^2}{n_p^2 - n_0^2} \quad (4)$$

with $F(r)$ a filling fraction calculated using the Maxwell-Garnett mixing rules, n_p the index of the aSi pillars and n_0 the index of the substrate.

We observe an overall equivalent transmission between the effective-index based and the hyperbolic metalens (See Fig. 9). However the difference is more visible for the reflection and the cross-talk. It can be explained by the final aspect of these metalenses and particularly their pillar distribution. The effective-index based metalens has less pillar due to phase-shift value out of reach with the existing index contrast. With less pillar on its trajectory the reflection decreases and the transmission increases but we observe also an increase of the cross-talk due a more restricted phase range.

Conclusion

We studied metalenses integrated on BSI CMOS pixel able to focus light and increase the amount of light reaching the photodiode. We demonstrated overall performance improvement using a compact library involving an higher height/diameter ratio nanopillars. Performance evaluation were also conducted to compare hyperbolic and quadratic phase profile. Finally design of a metalens based on Maxwell-Garnett analytic calculation of effective index have been performed. Demonstrators on CMOS image sensor are currently under fabrication and characterization will be presented in a future paper. The use of optimization algorithm was not considered in the presented design and it may be explored for next generation metalenses.
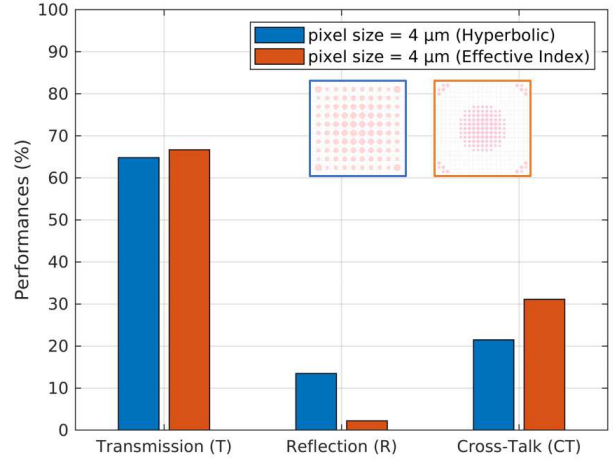
References

[1] L.Dilhan, J.Vaillant, A.Ostrovsky, L.Masarotto, C.Pichard, et R.Paquet, "Planar microlenses for near infrared CMOS image sensors", *Electron. Imaging, In vol.2020, n°7, p.144-1-144-7, janv.2020.*

[2] L.Dilhan *et al.*, "Planar microlenses applied to SPAD pixels", In: *IISW proceedings, 2021.*

[3] Steven J. Byrnes, Alan Lenef, Francesco Aieta, and Federico Capasso, "Designing large, high-efficiency, high-numerical-aperture, transmissive meta-lenses for visible light," In *Opt. Express 24, 5110-5124 (2016).*

[4] P.Lalanne, P.Chavel, "Metalenses at visible wavelengths: Past, present, perspectives.", In *Laser Photonics Rev. 2017, 11, 1600295.*

[5] P. Genevet, F. Capasso, F. Aieta, M. Khorasaninejad, and R. Devlin, "Recent advances in planar optics: from plasmonic to dielectric meta-surfaces", In *Optica 4, 139-152 (2017).*

[6] H.Benisty, J.Greffet, P.Lalanne, "Introduction to Nanophotonics", *Oxford Graduate Texts.*

[7] Augusto Martins *et al.*, "On Metalenses with Arbitrarily Wide Field of View", In: *ACS Photonics 2022 7 (8), 2073-2079.*

[8] E.Mikheeva *et al.*, "CMOS-compatible all-dielectric metalens for improving pixel photodetector arrays", In *APL Photonics 5, 116105 (2020).*

A hybrid, back-illuminated image sensor for high QE visible and infrared detection

A. Scott[a], P. Adamiec[c], S. Bednarski[c], A. Bofill-Petit[a], G. Di Nicolantonio[e], G. Kottaras[b], G. Margutti[e], L. Pancheri[d], A. Papathanasiou[b], A. Psomoulis[b], M. Sannino[a], E. Sarris[b], R. Turchetta[a]*, K. Minoglou[f]

a: IMASENIC S.L., Pl. Tetuan 40-41, 08010 Barcelona, Spain
b: SpaceASICS, Athens, Greece
c: Alter, Madrid, Spain, and Livingston, UK
d: TIFPA-INFN and University of Trento, Trento, Italy
e: LFoundry s.r.l., Avezzano, Italy
f: ESTEC, Nordvijk, The Netherlands

* Tel: +34 935466100   Email: renato.turchetta@imasenic.com

Abstract — For future space science and earth observation missions, the need for an image sensor with a high quantum efficiency over a wide band was identified. Having separate substrate for the detecting layer and the readout circuit allows for better optimization of the property of the substrate. The solution presented here is to have a thicker, back-side illuminated silicon die for the detector layer so that high QE can be achieved in a wide band, extending from visible to the near infrared.

In this paper we present the design of a 1Megapixel, BSI hybrid image sensor, which comprises of a detector and a CMOS readout integrated circuit (ROIC) layers. The detector was also manufactured in a modified CMOS image sensor technology. Test results of the detector and initial test of the ROIC will be presented.

I. INTRODUCTION

Most CMOS image sensors are fabricated on wafers with a thin epitaxial layer (~a few µm), which provides the sensing substrate. This selection matches well the requirements of vision applications, but it limits the quantum efficiency, especially for longer wavelengths. For space and earth observation applications, the specifications are driven by the need for high-efficiency over a wide band. This means it is possible to use a single sensor for many applications, thus limiting weight, power consumption and complexity, which are at premium in embarked instruments.

A hybrid detector (Figure 1) allow to separately optimise the sensing and the readout layers. For the former, the silicon photodetector array (SiPDA), we selected a 100 µm thick silicon substrate, while for the latter a CMOS readout integrated circuit (ROIC) was

developed. The two are connected together by bump bonding and can then be mounted in a standard package.

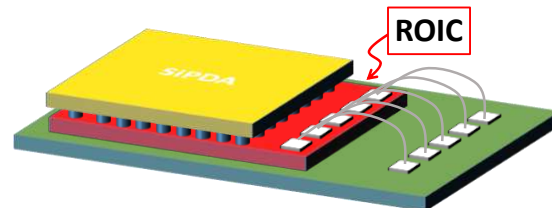The format of the hybrid array is 1040x1040 or 1 megapixel, with a pitch of 20 µm.

Figure 1. Artistic view of the hybrid sensor

II. THE SILICON PHOTODIODE ARRAY

The target process for the SiPDA is a modified 110nm CMOS (LF11IS) process [1]. As a substrate we used a low dark current, high resistivity (>2kOhm cm), float zone, 100 µm thick substrate with a backside p+ implantation and Anti-Reflective Coating. The mask set for the SiPDA is limited to only a few layers, which are going to define the implants for the pixel, together with contacts and one metal layer for the connections (Figure 3). Around the pixel array, guard rings are integrated and the outermost ring is used to bias the substrate: the bias voltage is transferred to the backside p+ layer through the non-depleted peripheral volume of the sensor chip. In this way, no back contact is needed. As shown in Figure 1, the SiPDA is slightly smaller than the ROIC, so all biases are provided through dedicated pads on the ROIC and routed to corresponding pads for bump bonding to the SiPDA. The guard-ring and the substrate contact geometry were defined by TCAD simulation to make smooth transition of the silicon

[1] L. Pancheri et al., A 110nm CMOS process with fully depleted high resistivity substrate for NIR, X-ray and charged particle imaging, Proceedings of the 2019 International mage Sensor Workshop, Snowbird Resort, USA, June 24-27, 2019

electric field from the pixels to the substrate contact. For the pixels, we consider different options for the N+ implants and the P+ isolations, as well as different geometries. In the simulation, we considered the parameters listed in Table 1. A simulation of the expected transmission and quantum efficiency for the sensor with the selected ARC and 100μm thickness is shown in Figure 4.

The megapixel array was divided in twelve subarrays, each with a variant of the design. In order to avoid breakdown of one structure to disrupt the correct functioning of their neighbours, guard-rings were inserted in between the subarrays. Stand-alone test structures were also designed to allow early characterisation of the pixel parameters. The manufactured wafer, together with a zoom to the reticle (Figure 5) and to the test structures (Figure 6) is shown in the referenced figures.

III. THE READOUT INTEGRATED CIRCUIT (ROIC)

A 150nm CMOS standard process (LF15A) was selected for the ROIC. The ROIC was partitioned in: an analogue ROIC (A-ROIC), comprising the pixel array with its row control as well as the programmable gain amplifiers (PGA) and a temperature sensor: a digital ROIC (D-ROIC) which includes the multi-column ADCs, followed by a piece-wise linearity correction block, the sequencer and a serial-to-parallel interface (SPI).

In the A-ROIC, the pixel analogue front-end (AFE) has to match 1-to-1 the pixel array in the SiPDA. The AFE has to be able to work in global shutter as well as provide dual gain, with the gain being selectable on a row-by-row basis.

The schematic of the pixel is shown in Figure 7. The diode is in the SiPDA and the line to the input of the source follower MSF corresponds to the hybridising bump. The capacitance C0 is used to adjust the gain of the pixel: when SWG is high, it is connected in parallel to the diode and the pixel works in low gain LG mode; while when SWG is pulled low,

C0 is disconnected and the diode works in high gain HG mode. The gain of the pixels can be selected in a row-by-row and frame-by-frame basis. An in-pixel bias transistor MB is provided to help with the global shutter operation. The command lines GR and GS control the shutter transistors MGR and MGS. When activated together with the bias VBIAS, the voltage at the output of the source follower is copied in the in-pixel sampling capacitance. The two samples for the signal and the reset are read in parallel on a dual analogue bus.
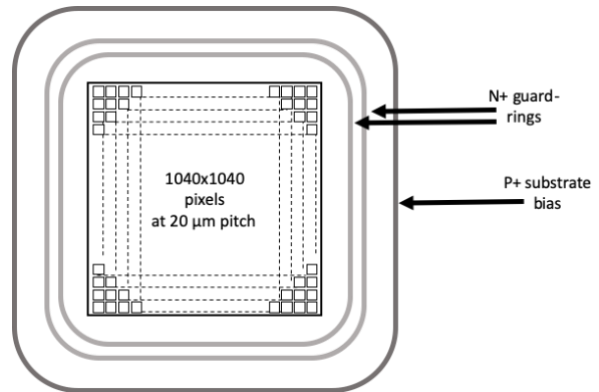
Figure 2. Layout of the SiPDA

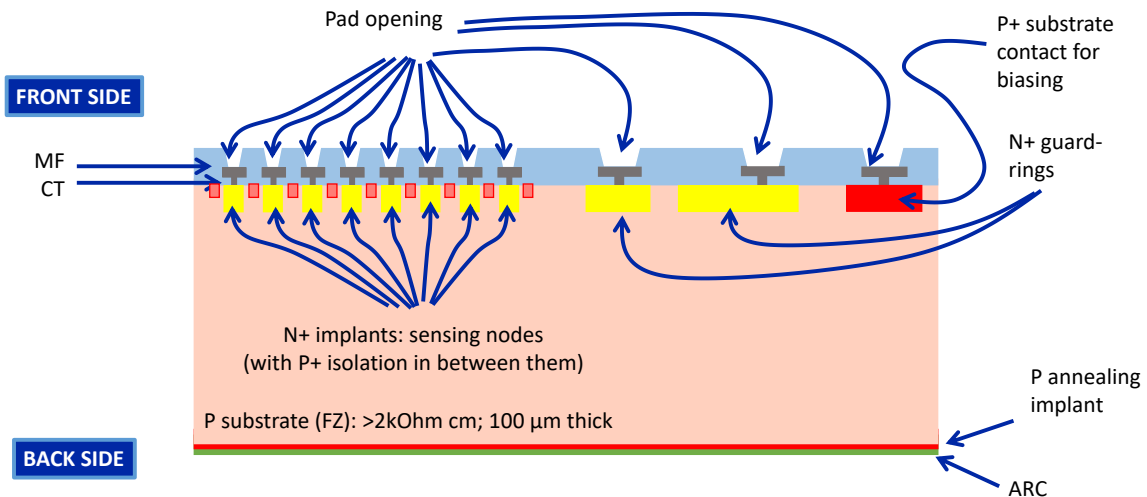| Parameter	Optimisation goal
Capacitance	Minimise
Surface leakage current	Minimise
Breakdown voltage	Need to be higher (in absolute value) than the depletion voltage
Pwell isolation bias voltage	Need to be lower (in absolute value) than the breakdown voltage
Full depletion voltage	Minimise

Table 1 Pixel figures of merit and criteria.

Figure 3. Cross-section of the SiPDA wafer. Drawing not in scale.

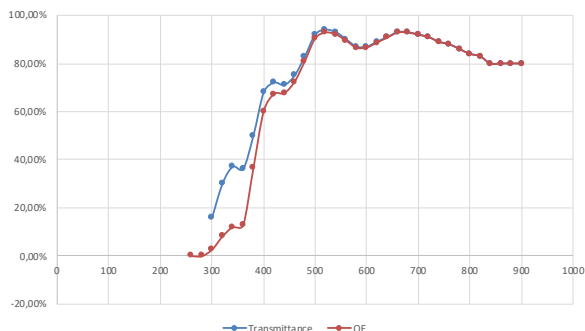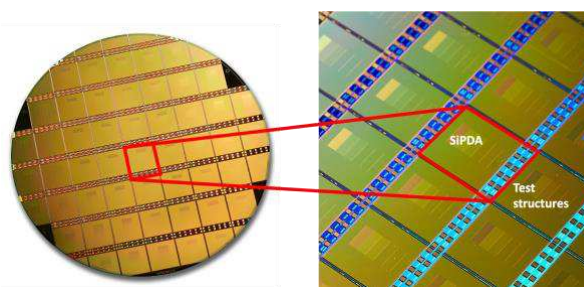*Figure 4. Simulated QE, showing the sensor achieves QE>75% over the 450 to 900 nm band.*

Figure 5. Photo of the SiPDA wafer with a zoom to the reticle

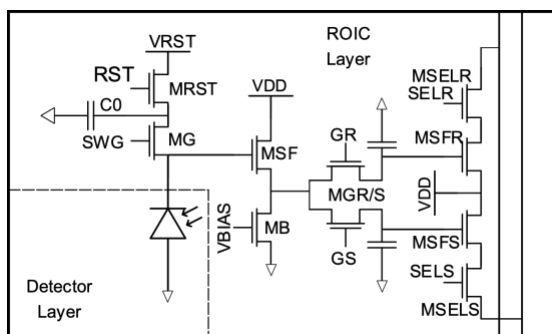*Figure 6. Photo of the test structures on the SiPDA wafer.*

Figure 7. Schematic of the GS pixel, with CDS. The gain is set on a row-by-row basis.

A few variants of the pixel were designed and integrated in the ROIC. The variants have slightly different analogue performance. The targeted operating temperature is 240K. At this temperature, the simulated noise is 12.1/102 e- rms in HG/LG with a corresponding full well of 102/1,110 ke-. At higher temperature of 293K, the noise increases slightly to 12.8/108 e rms and the full well does not change.
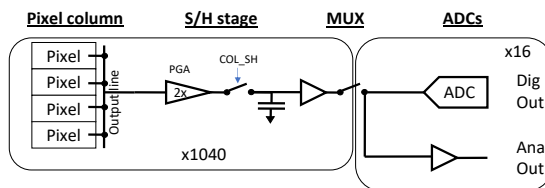
Figure 8. Schematic of the column analogue readout, including the ADC.

The pixel voltage is read through column-parallel programmable gain amplifier (PGA). Their gain can be set to x1, x2 or x4. In order to match the layout of the ADCs, after the sample-and-hold stage there is a 65:1 analogue multiplexer (Figure 9).
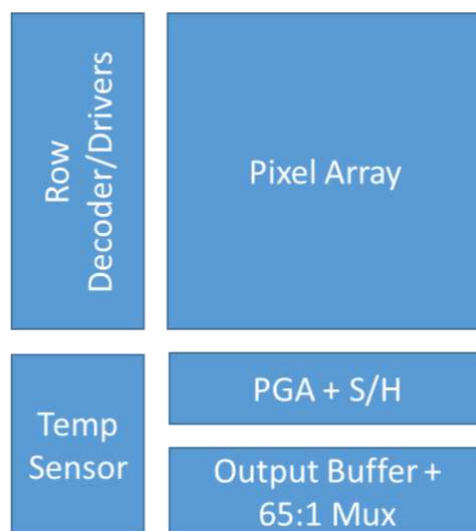
Figure 9. Floorplan of the A-ROIC.
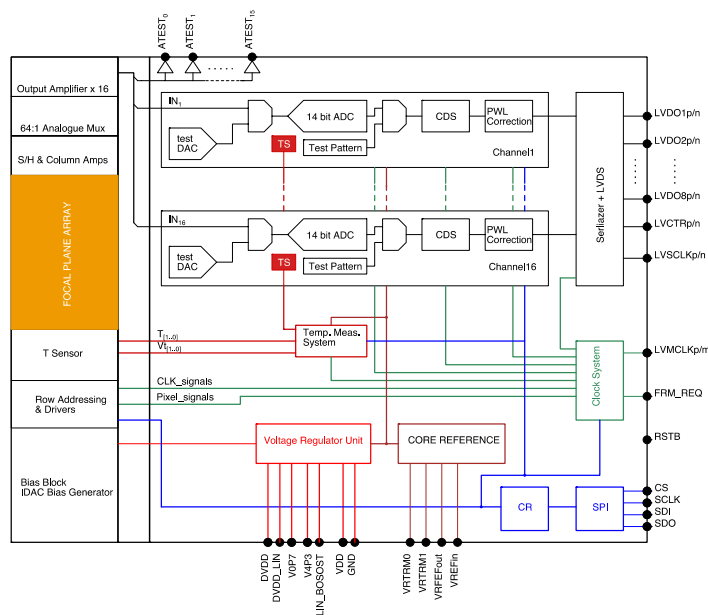
Figure 10. Floorplan of the D-ROIC

The ADC architecture is based on the hybrid successive approximation topology, with a 14-bit resolution. The conversion requires 10 clock cycles and with a clock of 100 MHz the conversion time is

100 nsec.

As a backup solution, an analogue output is also provided. The digitalisation occurs then off-chip. A photo of the packaged ROIC is shown in Figure 11.

Figure 11. Photo of the packaged ROIC.

IV. EXPERIMENTAL RESULTS

The test structures on the SiPDA provided early insight on the behaviour of the sensor before hybridisation. With respect to the parameters listed in Table 1, all but the inter-pixel isolation can be measured on the test structure. Each array in the test structure includes 60x60 pixels, all connected together. In Figure 12, the capacitance is measured for different test arrays as a function of the voltage. The line corresponding to the theoretical minimum capacitance corresponding to the full depletion of the 100 μm thick substrate is also shown. The curve indicates that full depletion is achieved in all test structures for a low bias voltage of around 25-30V.

Figure 13 shows the measured dark current as a function of the applied back-bias for the different test structures. A back-bias as high as 200V was applied, well above the full depletion voltage, and no breakdown was detected, showing that all pixel variants are safe to operate in full depletion.
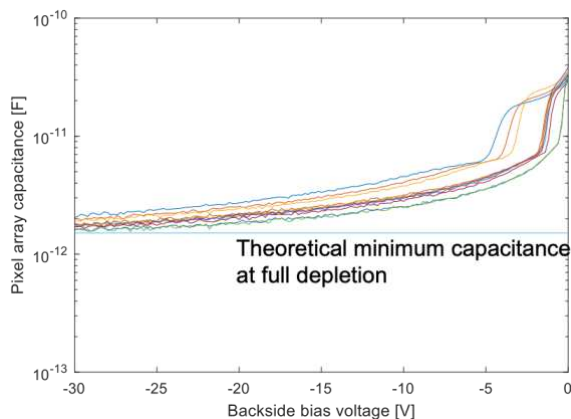
Figure 12. Measured capacitance from different types of pixels on the SiPDA test structure. It shows full depletion is achieved from around 20 V bias.
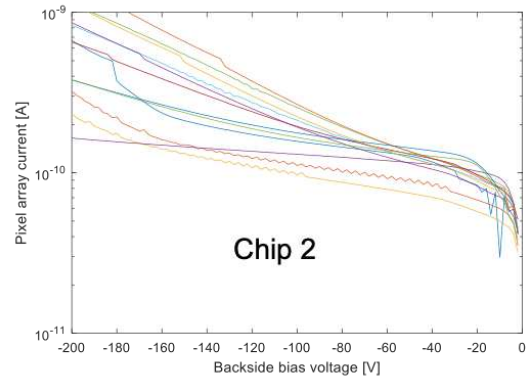
Figure 13. Measured dark current from different types of pixels on the SiPDA test structure.

Figure 14 summarises the dark current measurement for all 12 pixel variants and for 3 different chips. The pixel types can be ordered by implants (three groups: B00 to 03; B04 to 07; B08 to 11) and within each groups similar layouts are used. The measurements show some trends, with the first group (B00 to 03) and the third layout (B02, B06 and B10) having lower leakage. These results can be used for future optimisation of the SiPDA.
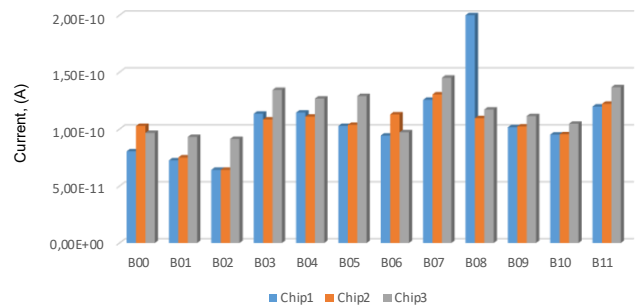

Figure 14. Measured dark current as a function of the design number B00 to B11. The colours refer to data measured on different chips. The dark current seems to be dominated by surface current generation. A new design with improved results is in preparation.

The ROIC is being currently tested before hybridisation. Most functionalities have already been proven. We expect the hybridisation to start soon and being under way by the time of the conference.

V. CONCLUSIONS

A megapixel, global shutter, high QE over a wide band hybrid image sensor was designed and manufactured. Initial results of the two layers, the SiPDA and the ROIC, taken individually show their good performance. The two layers should soon be sent for hybridisation with bump-bonding.

VI. ACKNOWLEDGEMENT

On-Chip Narrow Angle Filter Development

Dmitry Veinger
Tower Semiconductor
Shaul Amor 20, Migdal Haaemek, Israel
dmitry.veinger@towersemi.com

Naor Inbar
Tower Semiconductor
Shaul Amor 20, Migdal Haaemek, Israel
naorin@towersemi.com

Adi Birman
Tower Semiconductor
Shaul Amor 20, Migdal Haaemek, Israel
adibi@towersemi.com

Amos Fenigstein
Tower Semiconductor
Shaul Amor 20, Migdal Haaemek, Israel
amosfe@towersemi.com

Shirly Regev
Etesian Semiconductor Ltd
P.O.B 3227, Ramat Yishai, Israel
shirly.regev@etesiansemi.com

Dmitri Ivanov
Tower Semiconductor
Shaul Amor 20, Migdal Haaemek, Israel
dmitriiv@towersemi.com

Abstract — **This paper presents the narrow angle optical filter developed using a new process integration to enable the thick stack and the following pad opening**

Keywords — angle filter, thick optical stack, micro-lens array, process integration

I. Introduction

Certain imaging applications require an optical filter allowing only light propagating perpendicular to the sensor to be collected by their pixels. In recent years, one outstanding application for such a filter was in fingerprint sensor located under the cell phone's OLED screen [1]. For this application such a filter replaces the system lens which is often too thick to fit in modern thin cell phones. Another application example is a compound-eye camera [2].

The challenge is to implement such a filter directly on the Silicon for both a compact and cost-effective solution. To achieve this target, the required processing steps should be compatible with standard CMOS Image Sensor (CIS) capabilities. The main performance features of such a filter are the filter width, namely how fast the response drops for impinging beam angles larger than zero, and the quantum efficiency for the light coming in the right zero angle direction.

A straightforward approach to implement such a filter is by using micro-lenses focused on small apertures in an opaque material over the photo diodes in the pixel array, as depicted in Fig. 1 [1]. Angle beams will focus away from the aperture and will be blocked. With smaller light spot and aperture, filtering of only vertically impinging beams becomes more and more efficient. However, the spot size is limited by the wave nature of the light, thus a quality filter requires large, highly curved micro-lens elevated high above the photodiode. Simple analytical expression for this dependence is

$$\Delta\phi = \left[\left(\frac{d}{f}\right)^2 + \left(\frac{\lambda}{D}\right)^2\right]^{1/2}$$

where "d" and "D" are the diameters of aperture and the micro-lens, f is the micro-lens focal length and λ is the wavelength [2]. The aperture is optimized for avoiding light penetration at high angles and is realized using a set of apertures in several layers of the CIS metal layers (Fig. 2). The thick organic optical layers add a challenge for pad opening. A new process integration was developed to enable the thick stack and the following pad opening.
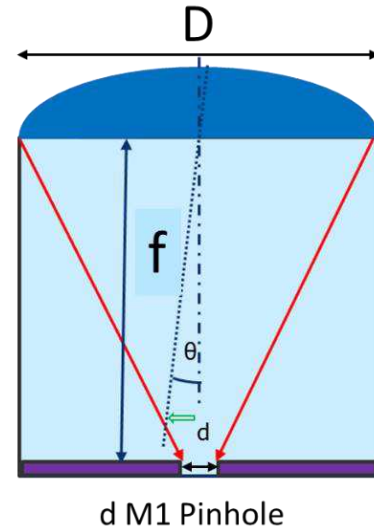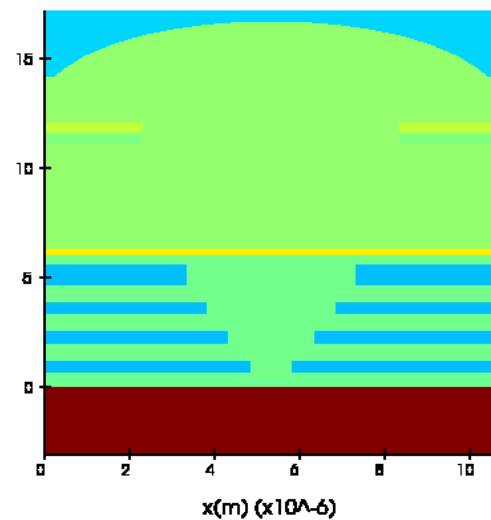

Fig. 1. – angle filter apparatus

Fig. 2. – Multilayer aperture implementation by metal (blue) and color filter layers

II. Filter Design

The micro-lens design was optimized by optical simulations. An example is shown in Fig. 3. The required angular sensitivity of the filter is defined by 90% signal drop at 5-degree illumination beam tilt. To satisfy the requirement, the bottom aperture diameter d was set to 1μm, the required

focal distance was 20μm, the micro-lens diameter D was checked in the range of 15-25μm and the micro-lens thickness (sag) was checked in the range of 5-9μm

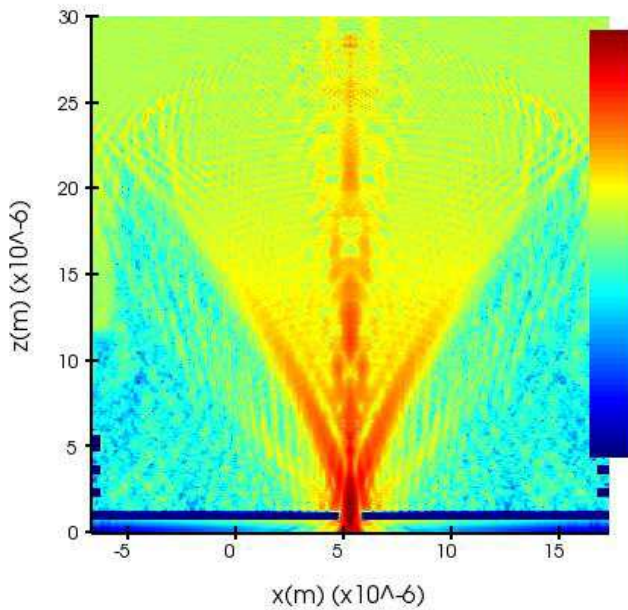Fig. 3. – optical simulation of the structure from Fig. 2: microlens diameter is 21μm, lens sag is 7μm and predicated focal distance is 20μm

III. THICK OPTICAL STACK INTEGRATION

The conventional micro-lens integration steps are described schematically in Fig.4. The process consists of the following steps:

1. Photo-imageable CT (transparent organic layer) depositions
2. CT exposure & development
3. Micro-lens (UL) coat on CT topography
4. UL exposure + development
5. UL melt

The narrow angle filter design requires much thicker CT layer than in the conventional optical back end. The thick organic optical stack introduces aggressive topography that puts a strong limitation on photo resist coating and patterning. Micro-lens photo-resist process over thick topography (5um or thicker) causes two main problems (Fig.5):

a. Photo-resist coating thickness non-uniformity
b. Photo-lithography focus issues (leveling)

The novel integration steps are described schematically in Fig.6. This integration enables a large range of thick stacks with relatively few process steps. An elegant approach was implemented where all steps of micro-lens formation were completed before the pad-opening topography is created. The new process consists of the following steps

1. CT coating
2. Flood (no mask) UV exposure for CT material bleaching
3. Micro-lens material coating and photo-lithography over planar surface
4. UL melting
5. Thin Low Temperature Oxide (LTO) layer deposition (used as a hard mask in the flow)

6. Photo-lithography to protect the lens array area
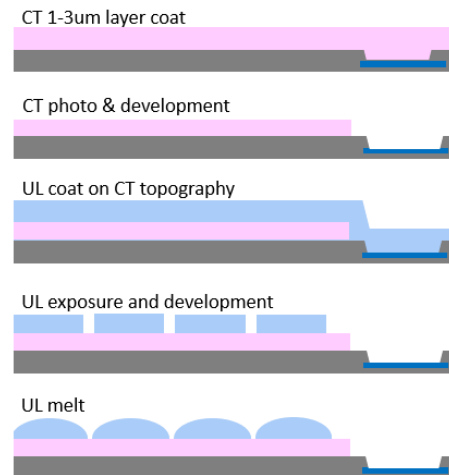7. LTO and thick CT layer etch

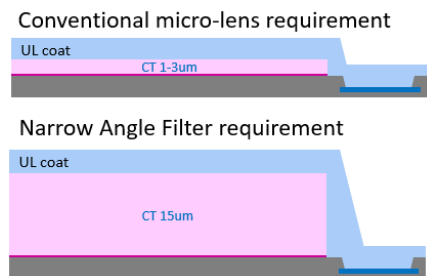
Fig. 4. – Conventional micro-lens process flow

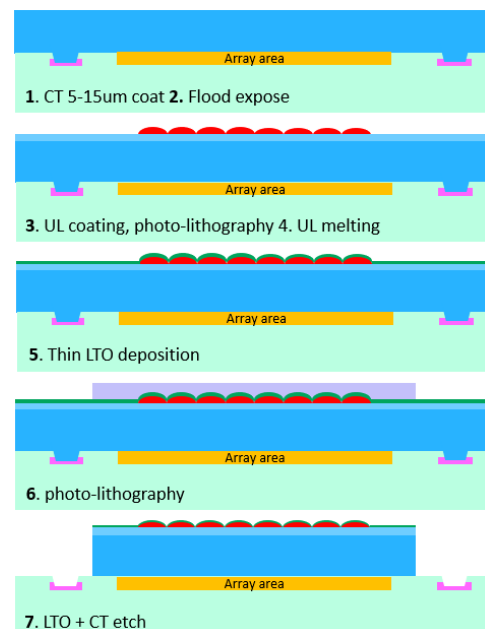Fig. 5. – Conventional micro-lens process flow over standard and very thick topography

Fig. 6. – the novel process flow for thick optical layers

Cross sections of the final optical stack shown are for the array (Fig. 7), array edge and pad area (Fig. 8)
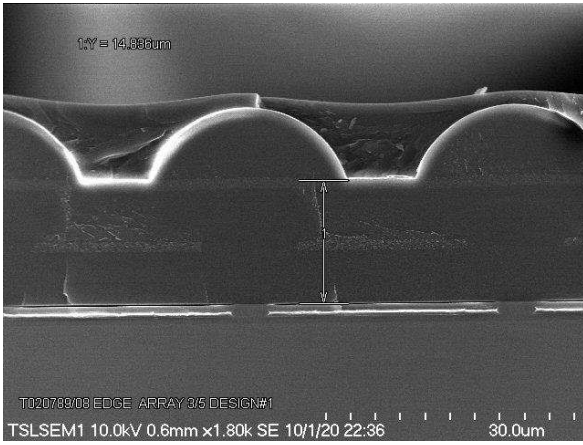
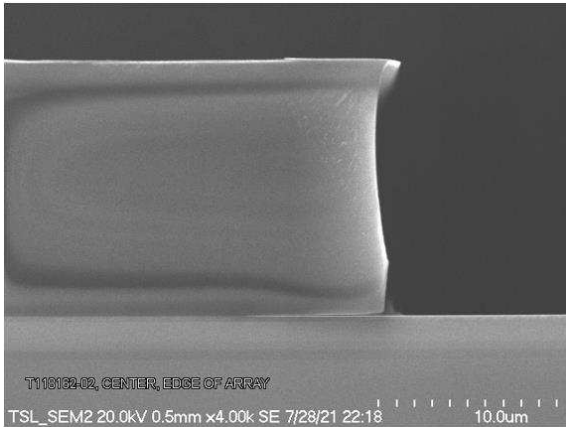Fig. 7. – Cross section of the optical stack over the array

Fig. 8. – Cross section of the optical stack on array edge near the pad opening area

IV. EXPERIMENTAL RESULTS

Results of empirical experiments with different apertures, lens sag, and elevation of 20μm are summarized in Table 1: Full Width Half Max (FWHM) of the filter and a stricter criterion for performance – angle where the normalized response drops to 10% of the maximum response at normal illumination.

TABLE I. SUMMARY OF 3 EXAMPLES OF OPTICAL STACK
OPTIMIZATION SHOWING PERFORMANCE OF THE FILTER

Lens Sag [um]	Micro-lens Diameter D [um]	Metal Aperture diameter d [um]	Color aperture diameter [um]	QE	Angle of 90% drop	FWHM angle
8	21.4	1	10	9.0%	5.2	2.7
7	21.4	1	10	8.8%	5.0	2.5
6	18.3	1	9	10.9%	5.2	2.8

Angle response curves and comparison to the simulated curve are shown in Fig. 9 and Fig. 10. The experimental results show good agreement with the simulations for 90% signal drop at 5-degree illumination tilt.
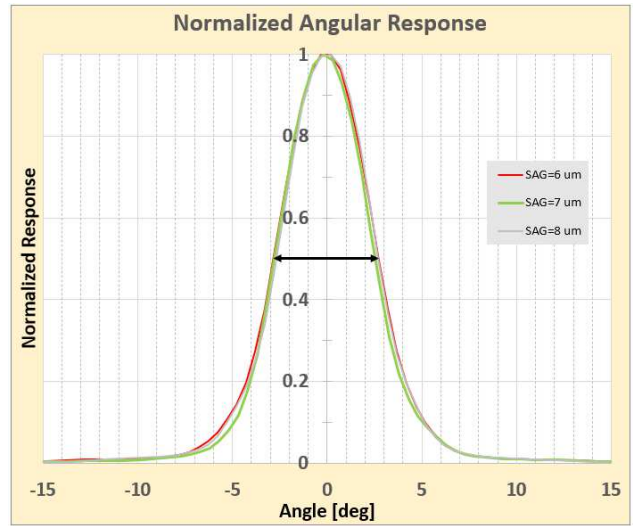

Fig. 9. – Measured filter performance, normalized response vs. impinging beam angle for the conditions from Table 1
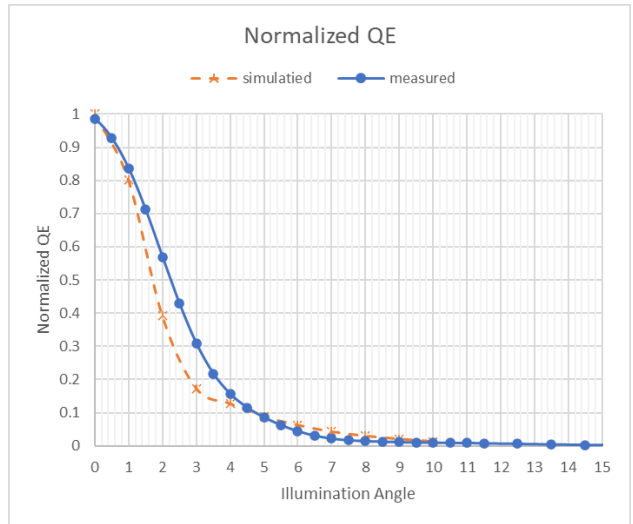

Fig. 10. – comparison of simulated filter performance to measured one for the 7 μm sag

REFERENCES

[1] Akkerman, H., Peeters, B., Van Breemen, A., Shanmugam, S., Ugalde Lopez, L., Tordera, D., van de Ketterij, R., Delvitto, E., Verbeek, R., Malinowski, P. and Ke, T.H., 2021. Integration of large‑area optical imagers for biometric recognition and touch in displays. Journal of the Society for Information Display, 29(12), pp.935-947.
[2] Duparré, J., Dannberg, P., Schreiber, P., Bräuer, A., & Tünnermann, A. (2005). Thin compound-eye camera. Applied optics, 44(15), 2949-2956.

Correlations between DCR and PDP of SPAD integrated in a 28 nm FD-SOI CMOS Technology

S. Gao [1], D. Issartel [1,3], M. Dolatpoor Lakeh [2], F. Mandorlo [1], R. Orobtchouk [1], J.-B. Kammerer [2], A. Cathelin [3], D. Golanski [3], W. Uhring [2], F. Calmon [1*]

[1] Univ Lyon, INSA Lyon, CNRS, Ecole Centrale de Lyon, Université Claude Bernard Lyon 1, CPE Lyon, INL, UMR5270, 69621 Villeurbanne, France
[2] ICube, University of Strasbourg, UMR CNRS 7357, Strasbourg, France
[3] STMicroelectronics, Crolles, France

Contact author: Francis Calmon, francis.calmon@insa-lyon.fr

Abstract — **This article presents an experimental study of the Dark Count Rate (DCR) and the Photon Detection Probability (PDP) of Single-Photon Avalanche Diodes (SPAD) implemented in 28nm Fully Depleted Silicon-On-Insulator (FD-SOI) CMOS technology and proposes a TCAD simulation study correlating the experimental DCR et PDP results. The measurements and the TCAD simulations show a dependence of DCR and PDP with the amount of the Shallow Trench Isolation (STI) oxide in the active zone and the alignment of the peripheral STI with the SPAD junction. SPADs featuring more STI oxide in the active zone (including the peripheral STI) tend to present higher DCR and PDP. The TCAD simulations show higher values of Avalanche Triggering Probability (ATP) explaining higher PDP obtained with SPAD featuring more STI oxide in the active zone. The defects located at the interface of P-Well / STI and related to the doping steps are potentially the one of the main sources of the higher DCR of the aforementioned SPAD.**

Key words — SPAD; 28nm FD-SOI CMOS; Dark Count Rate DCR; Photon Detection Probability PDP; Shallow Trench Isolation STI; TCAD Simulation; Avalanche Triggering Probability ATP

I. INTRODUCTION

Thanks to their high luminous sensitivity and sub-nanosecond response time, Single-Photon Avalanche Diodes (SPADs) are considered as one of the most suitable candidates in the field of photon counting [1], light detection and ranging (LIDAR) [2], and all applications requiring high sensitivity and fast response time. Also known as Geiger-mode Avalanche Diodes, SPADs have been intensively investigated both in academic research and industry. The Si-based SPADs have been already widely used for visible and near-infrared light due to their compatibility with the CMOS technologies [3][4], with remarkable performances for small SPAD pitches and/or large matrices [5][6][7]. As an

alternative architecture, SPAD devices have been successfully implemented in Ultra-Thin Body and BOX (UTBB) 28nm Fully Depleted Silicon-On-Insulator (FD-SOI) CMOS technology [8], in which the SPAD devices are integrated below the buried oxide layer (BOX) and the associated electronics is implemented in the ultra-thin silicon layer above the BOX layer. This integration allows intrinsic 3D stack at die level and therefore a greater fill-factor.

The performances of SPADs can be estimated by several criteria. Two of the most crucial ones are the Dark Count Rate (DCR) and the Photon Detection Probability (PDP). The former is defined as the number of undesired avalanche events per second and characterises the intrinsic noise of the SPAD device. The latter is used to evaluate the optical efficiency of the SPAD device and is defined as the probability of a received incident photon to generate an avalanche event.

In previous work, optimized architectures have been proposed to reduce DCR level of SPAD implemented in FD-SOI technology [9]. In this work, the proposed SPAD devices are characterised both in DCR and PDP. TCAD simulations are performed in order to correlate the experimental results of DCR and PDP measurements. The following sections are organised as follow: Section II introduces the proposed SPAD implemented in 28nm FD-SOI CMOS technology; Section III and IV respectively present the DCR and PDP measurements; Section V proposes an analyse of TCAD simulations allowing cross correlations between DCR and PDP; Section VI concludes the manuscript and gives some perspectives.

II. SPAD INTEGRATED IN CMOS FD-SOI TECHNOLOGY

The SPAD FD-SOI reference architecture is presented in Figure 1-a, the diode is made of a P-well / deep N-well (DNW) junction below the BOX layer

and the thin silicon layer (UTBB FD-SOI technology). The main advantage of this approach is to natively obtain a 3D SPAD pixel at die level considering back side illumination after die thinning, and associated electronics on top of the diode. Thanks to a Multi Project Wafer (MPW) run with standard CMOS FD-SOI 28nm standard process (except DNW implant), SPADs (25µm diameter, pseudo-circular shape) have been processed and characterized with 200kΩ integrated resistance as passive quenching and external readout [8][9]. Figure 1-b and 1-c represent two alternatives of SPAD FD-SOI architectures.

(a) Reference architecture

(b) Removal of STI trenches above the diode (called "Fusion")

(c) Peripheral STI aligned (#STI5) and fusion (called "Peripherical STI aligned and fusion")
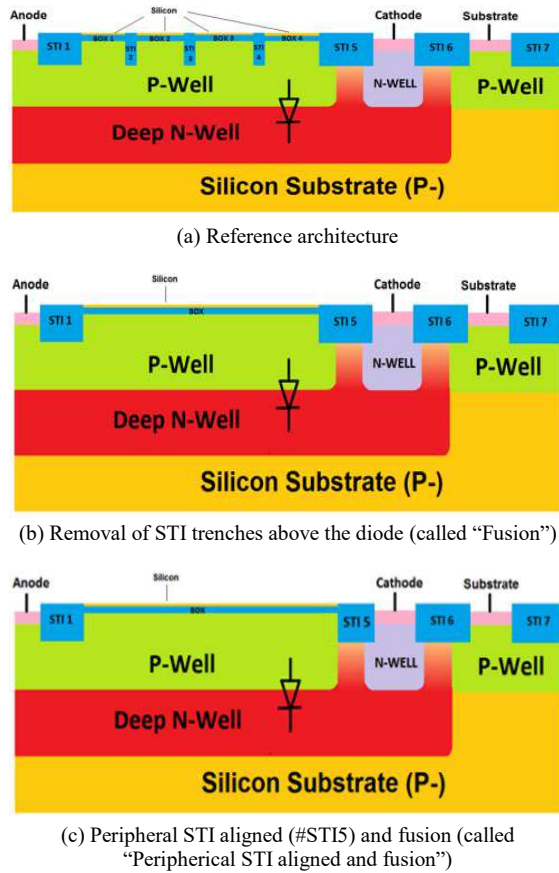
Figure 1: SPAD FDSOI schematic drawing of the three main studied structures with different positions of the STI regards to junction between P-Well and Deep N-Well.

In the reference architecture, design rules impose the presence of the Shallow Trench Isolation (STI) oxide in the active zone (represented by "STI 2", "STI 3" and "STI 4"). The defects located at the interfaces STI / P-Well are considered as the one of the main sources of DCR. The architecture called "Fusion" (Figure 1-b) has been proposed by removing (minimizing) these STI. Beyond the "Fusion" architecture, the peripheral STI ("STI 5") in the reference architecture is also considered as one major contribution to the DCR. The alignment of "STI 5" with peripheral junction (called "Peripheral STI aligned") alongside with the "Fusion"

architecture has been proposed to further reduce the DCR level (called "Peripheral STI aligned and fusion", Figure 1-c).

III. DCR STUDY

DCR has been intensively studied for specific architectures as depicted in Figure 1 (different STI locations) leading to different V_{ex} operating ranges above breakdown voltage around 15.8V at room temperature [9]. The STI locations in the active zone and at the periphery of the diode impact significantly the DCR performances (Figure 2). The "Fusion" architecture presents similar DCR level compared to reference architecture for V_{ex} below 7% V_{BD}. The alignment of peripheral STI allows to reach higher values of V_{ex}. A maximum excess voltage range of 20% can be reached for the architecture shown in Figure 1-c when removing the STI in the active zone and aligning the peripheral STI with the junction (i.e. structure called "Peripheral STI aligned and fusion"). This same architecture also presents the lowest DCR among the three aforementioned architectures for the same V_{ex}. Extracted activation energy, less than half the band gap of Si (between 0.2-0.3eV), indicates that the origin of the DCR can be a combination of band to band tunnelling and trap assisted or field enhanced generation recombination mechanisms.
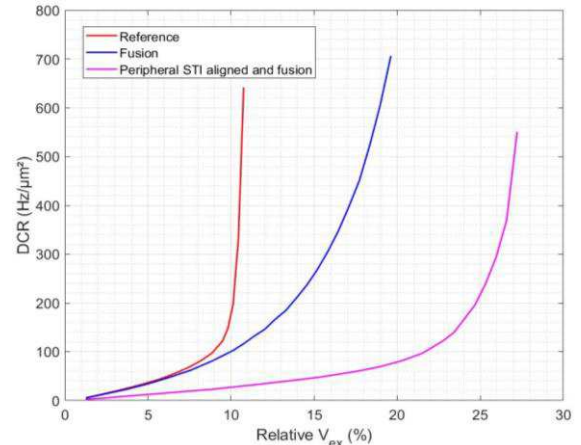

Figure 2: DCR versus relative excess voltage for the different architectures presented in Figure 1 (20°C)

IV. PDP STUDY

Measured PDP (front side illumination) for the different architectures are represented in Figure 3 (V_{ex} = 0.8V i.e. ~5%V_{BD}). The standard process is used (except DNW implant) including all back end of line layers (BEOL) without antireflective coating (ARC), micro-lens. The reference architecture presents the highest PDP, but the lowest performances in terms of DCR as maximum V_{ex} is around 10% (Figure 2). On the other hand, the "Peripheral STI aligned and fusion" architecture

presents the lowest values of PDP (PDP$_{max}$ is around 4.2% for V_{ex} = 0.8V at 620 nm of wavelength) while presenting the best DCR performances, since it can reach 20% higher values of V_{ex} and the reference architecture exhibits 2 to 4 times higher values of DCR. The performance of the "Fusion" architecture stands in between of the other two architectures, with PDP$_{max}$ being around 5% for V_{ex} = 0.8V at 620 nm of wavelength.

Then, we characterized the optimized architecture for DCR ("Peripheral STI aligned and fusion" Figure 1-c) varying V_{ex} up to 20% in Figure 4 with a maximum measured PDP around 12% at 620 nm of wavelength. This time, although the reference architecture exhibits higher PDP for the same V_{ex}, the "Peripheral STI aligned and fusion" architecture is more realistically advantageous due to the higher applicable V_{ex}. The PDP$_{max}$ around 12% is comparable to some studies in the literature [10][11][12].
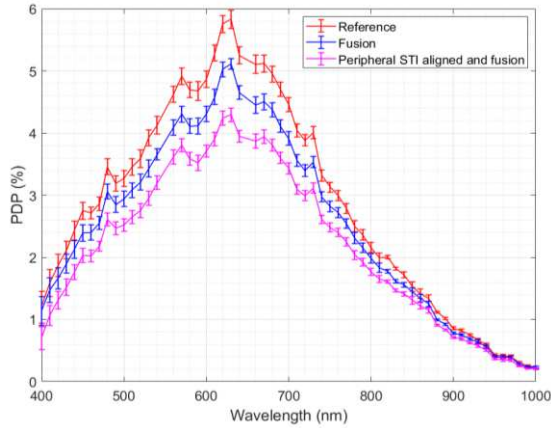
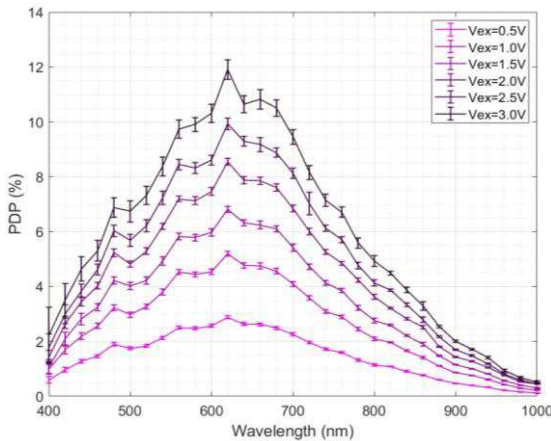Figure 3: PDP for the different architectures presented in Figure 1 (Vex = 0.8V, 20°C)

Figure 4: PDP for different V_{ex} (architecture: Peripheral STI aligned and fusion), 20°C

V. ANALYSIS AND COMPLEMENTARY TCAD SIMULATIONS

To further correlate the PDP and DCR measurements, the PDP$_{max}$ results have been plotted as a function of the DCR measurements in Figure 5.

The points {PDP$_{max}$-DCR} of the three architectures at same V_{ex} = 0.8V are highlighted in the triangle. The reference architecture, with the most amount of STI trenches (in the zone active and at the periphery) presents both higher DCR and PDP results, while the opposite case applies to the "Peripheral STI aligned and fusion" architecture. Now considering the points PDP$_{max}$ versus DCR with different V_{ex} for the "Peripheral STI aligned and fusion" architecture (green symbols), we observe a quasi-linear behaviour.
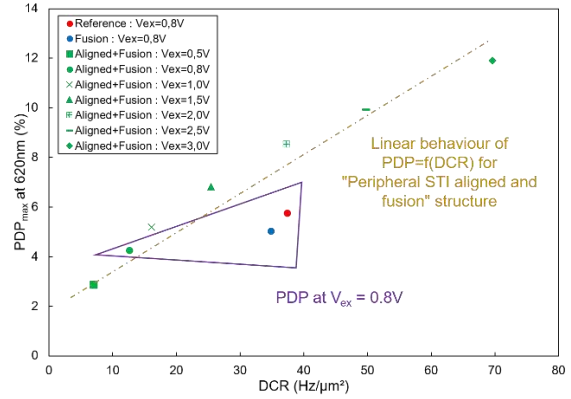

Figure 5: Maximum PDP at 620 nm of wavelength versus DCR for different structures and biasing voltages (20°C)

TCAD simulations indicate that the electric field is maximum at the diode periphery whatever the architecture. Moreover, for a given bias voltage, the reference architecture (Figure 1-a) exhibits the highest electric field and consequently the highest Avalanche Triggering Probability (ATP) compared to other architectures (observed in Figure 6). This remark explains the higher PDP observed for the "reference" structure compared to the "Peripheral STI aligned and fusion" one for the same excess voltage V_{ex}.

(a) Reference

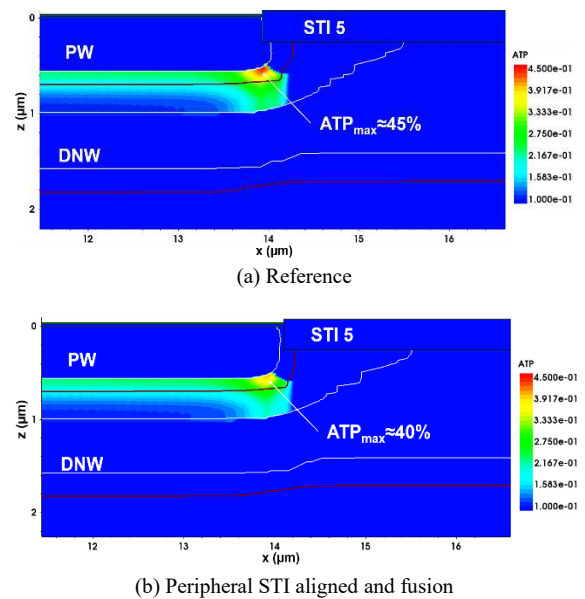(b) Peripheral STI aligned and fusion

Figure 6: Simulated Avalanche Triggering Probability (ATP) maps at the diode periphery

When including the defects located at the STI interfaces (that contribute to DCR), simulated dark current level decreases when the peripheral STI is shifted at the edge of the junction (architecture called "Peripheral STI aligned", Figure 7). Then we conclude that the "reference" structure presents the higher PDP due to higher electric field and ATP at the diode periphery, and the higher DCR due to STI defects.
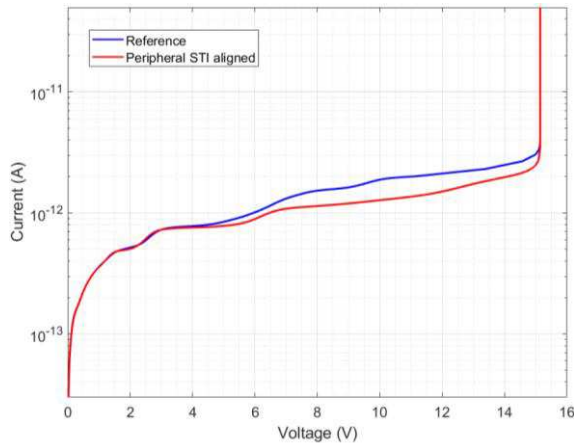

Figure 7: Simulated $I(V)$ curves with defects located at STI interfaces for the "reference" and the "Peripheral STI aligned" structures

VI. CONCLUSION AND PERSPECTIVES

Experimental DCR and PDP results obtained with passive quenching and external readout for SPAD integrated on CMOS FD-SOI technology are presented. The impact of STI locations (in the active zone and at the periphery) is analysed and TCAD simulations allows to understand the different experimental performances. Minimizing the amount of shallow trench isolation above the active region of the SPAD allows reducing the Dark Count Rate while extending the excess voltage range. Promising results have been obtained with a maximum PDP around 12% (at a wavelength of 620 nm) associated with a DCR less than 70 Hz/μm² (room temperature) at 20% excess bias voltage. Recently active quenching has been introduced and significant after-pulsing reduction has been demonstrated [13]. Ongoing work concerns PDP measurement for different STI patterning using approach introduced in [14].

REFERENCES

[1] M. Perenzoni et al. "A 160×120 Pixel Analog-Counting Singe-Photon Imager With Time-Gated and Self-Referenced Column-Parallel A/D Conversion for Fluorescence Lifetime Imaging," IEEE Journal of Solid-State Circuits, vol. 51, no. 1, pp. 155-167, 2015 (https://10.1109/JSSC.2015.2482497)

[2] I. Takai et al. "Single-Photon Avalanche Diode with Enhanced NIR-Sensitivity for Automotive LIDAR Systems," Sensors, vol. 16 no. 4, p. 459, 2016 (https://doi.org/10.3390/s16040459)

[3] W. Jiang et al. "Time-Gated and Multi-Junction SPADs in Standard 65nm CMOS Technology," IEEE Sensors Journal, t. 21, no. 10, p. 12092-12103, 2021 (https://doi.org/10.1109/JSEN.2021.3063319)

[4] S. Pellegrini et al. "Industrialised SPAD in 40nm technology," IEEE International Electron Devices Meeting (IEDM) 2017 (https://doi.org/10.1109/IEDM.2017.8268404)

[5] K. Morimoto et al. "3.2 Megapixel 3D-Stacked Charge Focusing SPAD for Low-Light Imaging and Depth Sensing," IEEE International Electron Devices Meeting (IEDM) 2021 (https://10.1109/IEDM19574.2021.9720605)

[6] S. Shimada et al. "A Back Illuminated 6µm SPAD Pixel Array with High PDE and Time Jitter Performance," IEEE International Electron Devices Meeting (IEDM) 2021 (https://10.1109/IEDM19574.2021.9720639)

[7] S. Shimada et al. "A SPAD Depth Sensor Robust Against Ambient Light: The Importance of Pixel Scaling and Demonstration of a 2.5µm Pixel with 21.8% PDE at 940nm, IEEE International Electron Devices Meeting (IEDM) 2022 (https:// 10.1109/IEDM45625.2022.10019414)

[8] T. Chaves de Albuquerque et al. "Integration of SPAD in 28nm FDSOI CMOS technology," ESSDERC 2018 (http://dx.doi.org/10.1109/ESSDERC.2018.8486852)

[9] D. Issartel et al. "Architecture optimization of SPAD integrated in 28 nm FD-SOI CMOS technology to reduce the DCR," Solid-State Electronics, Elsevier, Volume 191, April 2022, p. 108297 (https://doi.org/10.1016/j.sse.2022.108297)

[10] Z. You et al. "3µm Pitch, 1µm Active Diameter SPAD Arrays in 130nm CMOS Imaging Technology," IISW 2017, (Available: http://www.imagesensors.org/Past%20Workshops/2017%20Workshop/2017%20Papers/R21.pdf)

[11] A. Karz et al. "CMOS Single-Photon Avalanche Diode Pixel Design for a Gun Muzzle Flash Detection Camera," IEEE Transactions on Electron Devices, t. 65, no. 2, p. 547-554, 2018, (https://doi.org /10.1109/TED.2017.2779790)

[12] H. Ouh et al. "Combined In-Pixel Linear and Single-Photon Avalanche Diode Operation With Integrated Biasing for Wide-Dynamic-Range Optical Sensing," IEEE Journal of Solid-State Circuits, t. 55, no. 2, p. 392-403, 2020 (https://doi.org/10.1109/JSSC.2019.2944856)

[13] M. Dolatpoor Lakeh et al. "An Ultrafast Active Quenching Active Reset Circuit with 50 % SPAD Afterpulsing Reduction in a 28 nm FD-SOI CMOS Technology Using Body Biasing Technique" MDPI Sensors, 21(12), 4014, 2021 (https://doi.org/10.3390/s21124014)

[14] S. Gao et al. "3D Electro-optical Simulations for Improving the Photon Detection Probability of SPAD Implemented in FD-SOI CMOS Technology" SISPAD conference, 27-29 Sept. 2021 (https://doi.org/10.1109/SISPAD54002.2021.9592555)

A new digital pixel for particle detection

Nicola Massari
IRIS group
Fondazione Bruno Kessler
Via Sommarive 18,
Povo (TN), Italy
massari@fbk.eu

Luca Parmesan
IRIS group
Fondazione Bruno Kessler
Via Sommarive 18,
Povo (TN), Italy
luparmesan@fbk.eu

Gianluigi Casse
Department of Physics
University of Liverpool
Oxford Street, L69 7ZE,
Liverpool - UK
gcasse@liv.ac.uk

Abstract—In this work a new concept of binary pixel for particle detection has been introduced. The binary pixel, consisting of multiple junctions and coupled with an only-nmos circuit, tries to exploit the latchup effect to enhance the detection of particles and convert them directly into digital bit. A test array of 128x128 pixels has been realized in a 65nm standard CMOS technology showing a final pitch of 6.5 µm. Preliminary results demonstrate the capability of the sensor to detect both visible light (pulsed laser) and alpha particles (^{90}Sr).

Keywords—particle detectors, binary pixel, latchup effect

I. INTRODUCTION

Researchers want to study, in particle physics experiments, the nature of particles that constitute the matter and radiation in the universe. In these experiments, it is essential to use high performance custom detectors to enhance the detection capabilities of such systems. Nowadays the trend to build even more advanced experiments will result in the need for even more stringent requirements for the detector design in term of event rate and tracking capabilities [1]. It is desirable to design novel sensor architectures able to fast collect charges with a high number of pixels with reduced pitch to improve the position estimation of a particle hitting the active area. Moreover, power consumption has to be always minimized to ease the integration of multiple detectors in the system to cover the required area.

Most particle detectors found in the literature are based on a mixed-signal approach, having an analog front-end circuit, directly connected to the sensing node, followed by an analog to digital circuit, providing a multibit output to the readout electronics. The analog circuit has to amplify the useful signal, filter out possible noise contribution allowing the discrimination of the event. Due to the need to minimize mismatches and noise, the analog part of the design takes most of the area into the pixel. Typical pixel sizes for state-of-the art particle detectors range between 50×50 µm^2, in the case of hybrid pixel detector module pixels [2-3], and 28×28 µm^2 in Monolithic Active Pixel Sensors (MAPS) [4]. Moreover, the size of analog transistors, whose dimensions are driven by mismatch and noise requirements, do not scale with the same pace of digital transistor when more advanced technology nodes are used. For similar reasons, also power consumption of the analog front-end can be considered as an additional issue especially when the number of pixels of the array substantially increases. All these points prevent the realization of high-resolution detectors, representing a physical barrier for shrinking the pixel size towards µm or even sub-µm size. It is clear from this analysis that a new approach must be introduced to solve these issues.

In the proposed paper a novel concept of particle detector is presented. The main novelty of this approach consists of converting the acquired charge in digital signal as soon the event is detected. The binary pixel toggles when the particle has been detected, providing spatial information about the particle hit. This approach not only radically simplifies the pixel and the overall architecture, but also reduces the pixel size and the array power consumption. This fully digital concept was proposed already in [5] where a prototype based on 2 µm pitch with a bistable circuit was presented as a possible implementation. The pixel, thanks to the positive feedback, was able to toggle in a presence of less than 1000 e$^-$. Nevertheless, the collection efficiency was poor due to the use of standard substrate resistivity and of shallow junctions as sensing nodes. In this paper an exploratory pixel is shown. This new pixel exploits a positive feedback mechanism based on a controlled latch-up phenomenon that is triggered as soon a certain amount of charge is detected. Moreover, it uses a deeper junction as sensing node differently from [5]. The paper is organized as follows: the new pixel concept is presented in section II, while the array architecture in section III. Experimental results are eventually shown in section IV.

II. THE LATCH-UP PIXEL

The interaction of ionizing particles with digital circuits implemented in silicon gives rise to the observation of some unwanted behavior, called Single Event Upset (SEU), causing the change of state of some elements in the logic. Depending on the type of interaction, these events can be temporary or permanent, sometimes causing relevant current discharge and circuit damage.
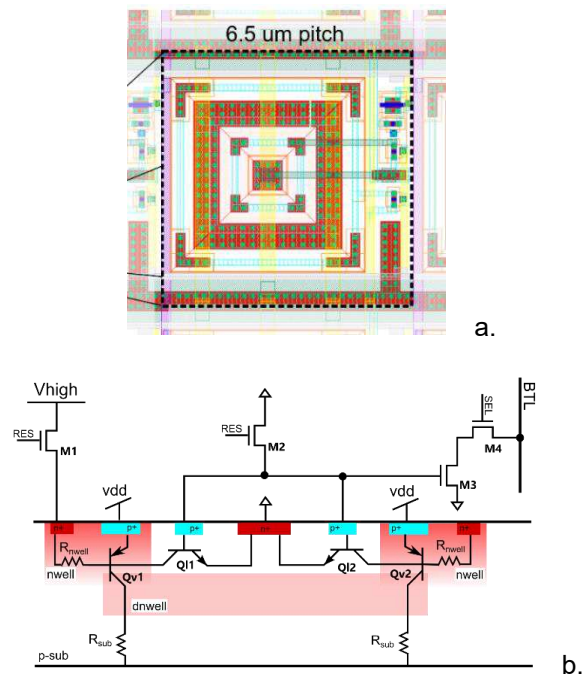

Figure 1: Cross-section (a.) and schematic (b.) of the pixel based on latch-up phenomenon.

A fully digital particle detector can be thought as an array of memory elements whose state, initialized at a specific value, may toggle when a particle hits the pixel. With this idea in mind, the new pixel is designed to be sensitive to the acquired charges favoring the change of status as soon as the sensitive node is stimulated by the particle transition.

The layout and cross-section of the pixel is shown in Figure 1a-b. As observed, the main collecting junction consists of a squared deep-n well connected through a n-well to the reference voltage V_{high}. Due to layout constraints, the pixel pitch is limited at 6.5μm, consisting of a minimum size deep-n well surrounded by similar deep-n wells placed at the minimum allowed distance. In the inner part of the pixel a central n+ junction is enclosed by a p+ ring, composing respectively the emitter and base of a lateral NPN BJT transistor. An only n-MOS circuit has been designed outside the sensing node to avoid the implementation of additional n-well, that may compete on the charge collection with the main junction. Next to the lateral BJT, other parasitic transistors can be identified. In particular, a vertical PNP BJT is formed by a p+ junction (emitter) embedded in the n well and forced to Vdd, by the n well (base) and the p- substrate (collector). In the cross-section, parasitic resistances R_{nwell} and R_{sub}, respectively referred to the n well and substrate, are also visible.
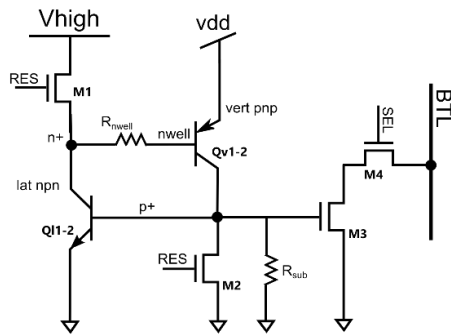

Figure 2: Schematic of the proposed pixel.

Figure 2 reports the schematic of the pixel including implemented MOS transistors and presence of active and passive parasitic devices for a better understanding of the pixel behavior. In this schematic transistors M_1 and M_2 are only used to pre-charge nodes n+ and p+ respectively to a reference value V_{high} and ground. The initial value of V_{high} is properly set to maximize the sensitivity of the pixel to the variation of node n+. After this first reset phase, both junctions are left floating, then free to move as soon as a particle is detected. The main junction (deep-n well) is connected through the n+ node and M_1 to V_{high} and when charges are generated and collected due to a particle hit, the voltage of node n+ suddenly decreases activating the vertical bjt Q_{v1-2}. The current flowing through Q_{v1-2} charges node p+ that activates the lateral npn transistor Q_{l1-2} thus implementing a positive feedback that toggles the status of the pixel. The pixel status will be readout by switching on transistor M4 (SEL=H) and using the common source amplifier M3 directly providing a binary value on the bit-line (BTL).

III. SENSOR ARCHITECTURE

The sensor array, designed in a 65nm technology, consists of 128x128 pixels. As shown in Figure 3, the array contains three different pixel topologies in order to compare their behavior when exposed to ionizing particles.
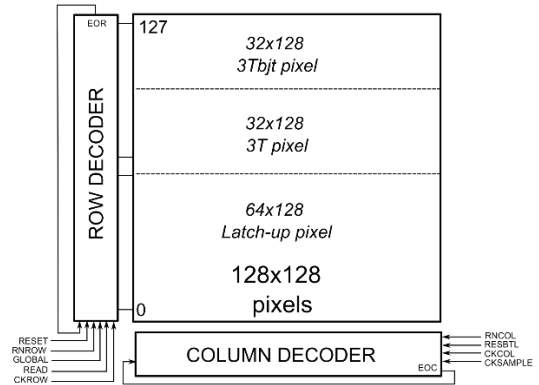

Figure 3: Architecture of the 128x128 pixels.

The detector is designed to work in rolling shutter mode using a row decoder for scanning all rows of the array sequentially and a column decoder for sampling data from the bit-lines and delivering the pixel content out of the chip. The readout consists of a register that, once sampled input values, serially scans data organized in byte (in this way reading an entire row needs 16 periods of clocks CKROW). Figure 4 shows the typical timing diagram needed for sensor initialization and data readout.
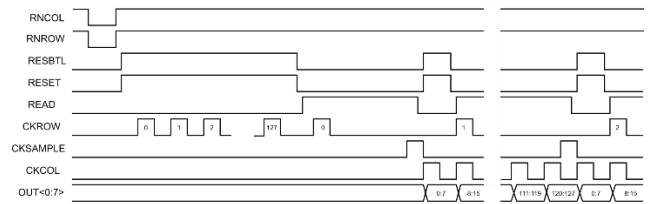

Figure 4: Timing diagram for pixel reset and readout phase of the array working in rolling shutter mode.

As observed in Figure 3, half of the array (64x128 pixels) is based on latch-up pixels. The remaining half array is divided in two other sub-arrays. The first sub-array of 32x128 pixels is made of a single deep-nwell junction readout using three transistors' pixels (3T pixel of Figure 3) while the last sub-array consists of 32x128 pixels based on vertical BJT (phototransistor, 3Tbjt of Figure 3).
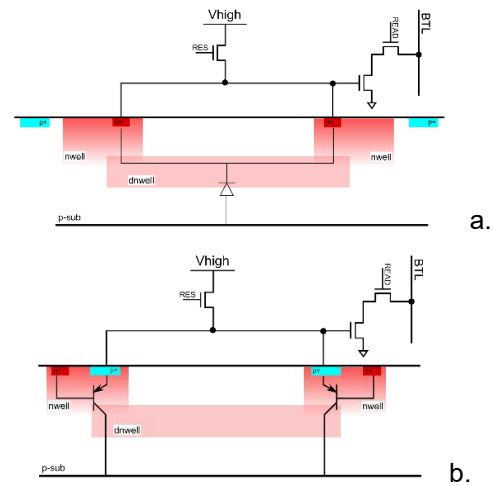

Figure 5: Cross section of 3T and bjt3T pixel of the proposed sensor array.

Figure 5 shows the cross section of the 3T and bjt3T pixel, having same size of the latch-up pixel. As observed, the 3T pixel mainly differs from the bjt3T pixel using an extra junction for the implementation of a vertical bipolar transistor. The goal in this case is to introduce an intrinsic gain in the charge collection operation.

IV. EXPERIMENTAL RESULTS

The characterization of the latch-up pixel was first done using a test structure external to the array. This structure allows us to have access to the analog value of the pixel output, being not possible to read it from the digital array. The test structure schematic in Figure 6, which shows the presence of electrical stimulator consisting of a MOS transistor working as a capacitance placed at the input of the pixel. After the pixel reset, a negative edge of amplitude V_{IN} is applied to the capacitance to emulate a certain amount of charge injected by the particle. Due to a problem in the design, only a maximum voltage can be applied (Vin = 1.2V), corresponding to an estimated value of 7000 e-. With this stimulus, the pixel is always able to toggle.
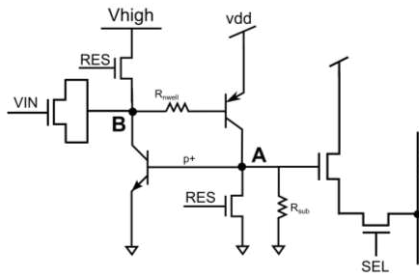

Figure 6: Schematic of the test structure: the input of the test pixel is connected through a MOS capacitance to an input voltage V_{IN} able to inject a certain amount of charge into the sensitive node.
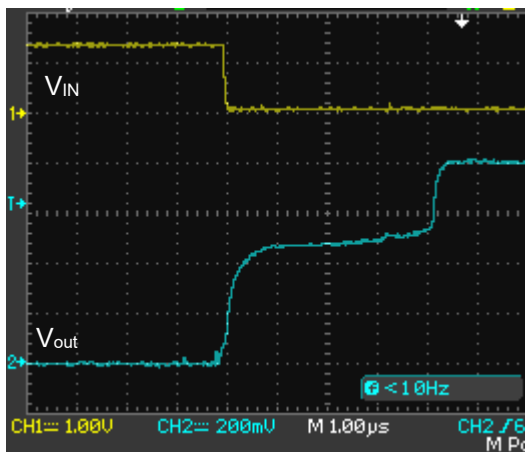

Figure 7: Analog value of the output of the test pixel acquired with an oscilloscope when an electrical stimulus is injected at the input.

Figure 7 shows the typical evolution of the output in correspondence of the stimulus. As observed, the injection of charge increases the value of node V_A until the positive feedback of the latch-up is activated (second rising edge), thus forcing the output voltage to Vdd (1.2V). The activation of the latch-up happens when the lateral BJT is enabled (V_A~0.7V).

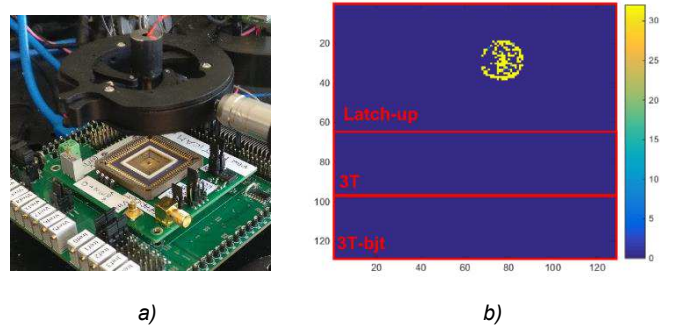
a)   b)

Figure 8: a) Measurement setup: a pulsed blue laser is focused on the active area of the array to check the response of the pixel to an optical stimulus; b) output of the array after the accumulation of 32 consecutive frames.

The array of pixels was also optically tested using visible light. In the setup of Figure 8a an external pulsed laser peaked at 490 nm was focused on the detector in order to stimulate the pixel with a certain amount of charge. The detector, synchronized with the source of light, acquired a sequence of subsequent 32 binary images that were accumulated to obtain a final image with 5b resolution (see Figure 8b). A burst of four subsequent laser pulses, impinging in the array, were synchronized at every image. As result, the projection of the laser spot is clearly visible while no activity is also observed in all other pixels.

The detector was also exposed to radiation to prove its capability to detect high energy particles. In this first experiment a ^{90}Sr was placed on top of the detector at minimum distance in order to maximize the flux of events. The image of Figure 9 is the results of the accumulation of 10000 frames of 35ms each. As observed, the image is clearly divided in two sub-array showing a different response to the exposure. The sub-array placed at the left of the image refers to the 3T and bjt3T pixels while the right part refers to the latch-up pixel. The distribution of events accumulated in every pixel after multiple exposures is clearly more uniform in the sub-array of 3T and bjt3T pixels than the sub-array of latch-up pixels. Nevertheless, the latch-up array seems to have, as average, pixels with more sensitivity than the other sub-array.
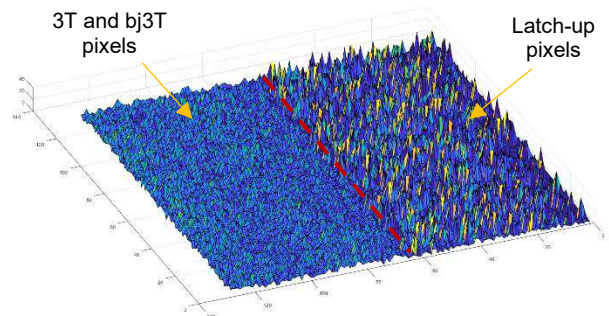

Figure 9: Image obtained after exposure to a ^{90}Sr source. The radioactive source was placed as close as possible to the active area to maximize the detection activity. Two main regions can be distinguished.

Figure 10 shows the histogram of number of events has been detected in al 10000 frame of the experiment for the two different sub-arrays. As result of this analysis, the rate of

detection in the latch-up array is higher than the 3T and bjt3T approach. The average number of detected particles in the latch-up array is equal to about 15, corresponding to an equivalent flux of about 480 events/s, while the other sub-array the rate is of about 280 events/s.
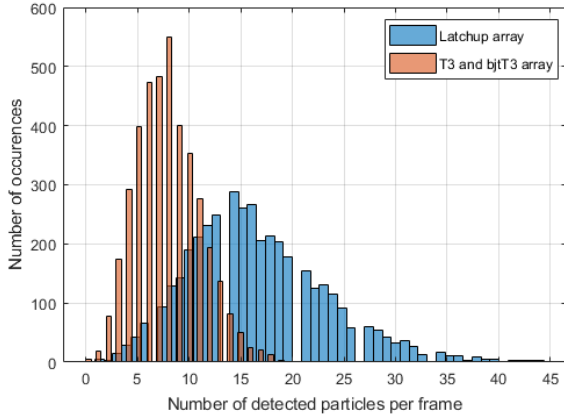

Figure 10: Number of detected events at different frames for the two sub-arrays.

Figure 11a shows the pixel value distribution of both 3T and bjt3T implementation. Such distribution can be approximated by a Poisson distribution with average value equal to 9.8 events per pixel. Figure 11b instead shows the distribution of latch-up pixel values. This histogram shows the overlap of multiple distributions with different coefficients due to pixel sensitivity variations across the array. We suppose the non-uniformity of the latch-up pixel is likely due to the variation of parameters of parasitic elements across the array that change the sensitivity of each pixel with respect to similar stimulus.
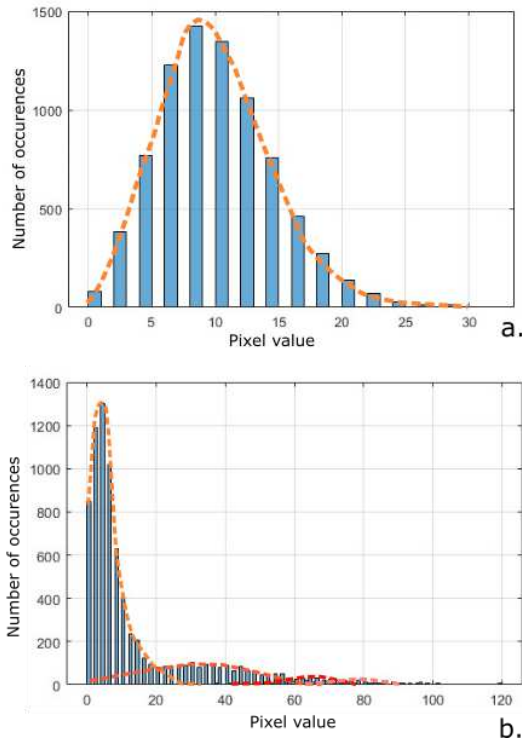

Figure 11: Distribution of pixel values in the two sub-arrays. a) distribution in the 3T and bjt3T pixel: a single distribution is clearly

visible. b) distribution of values in the latch-up pixel: multiple overlapped distributions are identified.
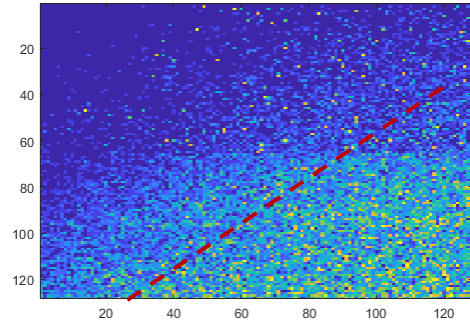

Figure 12: Events detected when the detector, exposed to ^{90}Sr source, is partially shielded by a sheet interposed between the source and the sensor.

The second experiment shows the response of the sensor with the same alpha source placed on the top of the detector and a sheet which partially cover the active area. As a result of this, only a portion of the array shows activity. In this experiment the different response of the two pixels topology is even more enhanced than the previous experiment.

V. CONCLUSIONS

The present paper shows the implementation of new fully digital pixel for particle detection. Preliminary results show that the detector responds to an optical stimulus as well as to a radioactive source. Images obtained after exposure to an alpha source (^{90}Sr) clearly show a spread of the pixel sensitivity in the array resulting in a relevant non-uniformity especially if compared with the image of another pixel topology (3T and bjt3T) implemented in the same array.

REFERENCES

[1] F. Hartmann, Evolution of Silicon Sensor Technology in Particle Physics 2nd edn (Springer, 2017).

[2] M. Garcia-Sciveres, N. Wermes, A review of advances in pixel detectors for experiments with high rate and radiation, Rep. Prog. Phys. 81, 066101 (2018)

[3] The RD53 collaboration, www.cern.ch/RD53.

[4] M. Mager, ALPIDE, the Monolithic Active Pixel Sensor for the ALICE ITS upgrade, Nucl. Instrum. Meth. A824 (2016) 434–438.

[5] G. Casse et al 2022 JINST 17 P04010.

A Study on a Feature Extractable CMOS Image Sensor for Low-Power Image Classification System

Shunsuke Okura, Ai Otani, Koshiro Itsuki, Yusuke Kitazawa,
Kohei Yamamoto, Yu Osuka, Yudai Morikaku, and Kota Yoshida
Ritsumeikan Univ., Shiga, Japan, sokura@fc.ritsumei.ac.jp

I. INTRODUCTION

In development of internet of things (IoT) with trillion sensor universe, the amount of information collected by CMOS image sensors will be drastically increased, and an image recognition based on deep learning (DL) is getting more important to process the big imaging data. However, the data captured by conventional image sensors is redundant for the DL because features of the imaging data is extracted in the deep neural networks (DNNs). For the feature extraction, convolution multiply-accumulate (MAC) operation takes place in the DNNs. Yoneda et al. has proposed an image sensor capable of analog convolution [1], in which convolution MAC operation is conducted in a pixel array. However, crystalline oxide semiconductor FET with an in-pixel capacitor are utilized to suppress leakage current during the convolution, thus resulting in large pixel size. Besides, operational transconductance amplifier (OTA) is utilized for current-domain MAC operation with larger power consumption compared to normal imaging mode. Young et al. has proposed a log-gradient image sensor [2], in which low-complexity feature for machine learning (ML), that is histogram of oriented gradients (HoG), is derived in a column readout circuit with analog four-line-memory. The HoG is aggressively quantized with a 2.75 bit ratio-to-digital converter (RDC), thus conventional RGB color image cannot be readout with the signal chain.

In this paper, a CMOS image sensor which can generate both normal image for human and feature data for the DL is proposed to reduce the power consumption of the image classification system and to save storage space for the big data. In order to keep compatibility with conventional image sensors, the CIS does not employ analog memory for convolution. Simulation results of image classification with horizontal edge as feature data of CMOS image sensor output is shown Sec. II. The CMOS image sensor which can generate the horizontal edge is presented in Sec. III, followed by summary and future work in Sec. IV.

II. IMAGE CLASSIFICATION SYSTEM WITH A FEATURE EXTRACTABLE CIS

Figure 1 shows a concept overview of our image classification system with the feature extractable CIS. The CIS is switched to an imaging mode according to a trigger generated by a CNN that detects person and/or other objects with the

feature data. At the feature mode, the power consumed by the CIS can be reduced with aggressive quantization such as 3 bit. Besides, the power consumption of the CNN will be reduced by omitting redundant layers and filter channels in the feature extractor designed for normal images.

In order to verify image classification accuracy with the feature data, the RGB color INRIA person dataset [3] is converted to a horizontal edge dataset and then input to a 3-layer CNN as shown in Fig. 2. The process to generate the horizontal edge dataset simulates the operation of the proposed CIS. The horizontal edge is difference between vertically adjacent pixels derived with y-derivative but without convolution, thus the 64×128 pixel input image is scaled down to 64×64 pixel. The noise is also added because pixel reset noise is not cancelled in the y-derivative as described in Sec. III. The quantizer represents low-bit analog-to-digital (A/D) conversion. The quantized horizontal edge is resized back to 64×128 pixels in order to use same size CNN for the comparison of original RGB dataset and the horizontal edge dataset. It is expensive to construct a new dataset with our feature-extractable CIS for training CNN models, but the cost can be almost negligible by transforming public datasets according to the behavior of our sensors. Image classification accuracy is summarized in Table I, in which 8 bit RGB color image at the imaging mode and 8 bit horizontal edge at the feature mode are utilized for the training and test of the CNN. The accuracy is 98.3% when the RGB color test dataset are classified with the CNN trained with RGB color image (1), while the accuracy drops to 47.2% with the edge dataset for test (2). The classification accuracy of the edge dataset is improved with the CNN trained with the edge dataset (3). The accuracy further increases to 97.0% when contrast of the edge data is enhanced with histogram equalization (4), which is comparable to the classification accuracy of original RGB color dataset. Figure 3 also shows simulation results of the image classification according to the size of the dataset with quantization. The horizontal edge is robust to the quantization down to 3 bit. Even though the accuracy decreases only by 1.4%, the size of the dataset decreases by 95% with the 3 bit edge dataset compared to the original 8 bit RGB color dataset.

This simulation experiments suggest that the ouput of a CMOS image sensor can be (a) horizontal edge without convolution and (b) low bit-resolution such as 3 bit, for image classification. The feature extractable CIS with the pixel capable of horizontal edge detection and the variable bit-

resolution successive approximation register (SAR) analog-to-digital converter (ADC) is described in the following section.

III. Column signal chain of the feature extractable CIS

Figure 4 shows a column signal chain of the feature extractable CIS that consists of pixels and a SAR-ADC. The pixel configuration is same as a conventional DCG pixel composed of 4T pixel and a binning gate BIN. The BIN gate is turned-on during the feature mode to derive the difference between vertically adjacent pixels for the y-derivative. The pixel source follower transistor SF is pre-charged with ΦPC [4] inactivating the current source I_d to save power consumption during the feature mode. The SAR-ADC is composed of a comparator CMP and a SAR digital-to-analog converter (DAC). A double clamp circuit [5] and the bias current for the latch in the comparator are inactivated during the feature mode also to save power consumption.

Figure 5 shows a timing diagram of the proposed CIS. At the imaging mode, the operation timing is same as a conventional 4T pixel, and the difference between the pixel reset and signal is converted into 10 bit digital code. At the feature mode, the signal levels of J-th and (J+1)-th row pixels are readout instead of the reset and signal levels of a selected pixel, so that the FD reset noise is not cancelled. However, the reset noise is acceptable because the feature data can be classified with the CNN at low bit-resolution such as 3 bit. The ADC converts the signal difference between J-th and (J+1)-th row pixels, that ranges from negative to positive, into 5 bit digital code with margin. The 9-th bit switch of the SAR-DAC is connected to V_{RL} and other switches are connected to V_{RH} during the sampling of J-th row pixel, thus 0.5 V offset voltage is added to the DAC output V_{DAC} prior to the A/D conversion. Then, 5 MSBs of SAR-DAC are switched during the A/D conversion. According to SPICE simulations, the current dissipation I_{dis} during the feature mode is only 0.317 [μA] that is reduced by 99.0% compared to that during the imaging mode as summarized in Table II.

The settling error of the pixel SF at the feature mode is a concern for large capacitive load on the pixel output column line due to weak inversion operation. However, the settling error that depends on the input signal level can be divided into gain error and linearity error after the double sampling of J-th and (J+1)-th row pixels, and the gain error affects less to the feature data. SPICE simulation results of settling error is shown in Fig 6, in which signal level of J-th and (J+1)-th row pixels are respectively swept from 0.0 V to 0.5 V thus the input difference of the SF input is swept from -0.5 V to $+0.5$ V. The gain of the pixel SF is given by 0.91, and the maximum linearity error is given by 7.26 mV which corresponds to 7.1 bit resolution and is acceptable for 5 bit A/D conversion.

A test chip of the variable SAR-ADC was implemented with 0.18 μm CMOS process [6]. Figure 7 shows the chip photograph of the 640 column ADC. Figure 8 shows the sample images for a ramp input signal. It is noted that the nonlinearity is caused by capacitance error of the split

capacitor C_{SP} in the SAR-DAC. Except for the split capacitor error, the DNL was 1.51 LSB at the 10 bit imaging mode and 0.12 LSB at the 5 bit feature mode. At the imaging mode, monotonicity and small column FPN were confirmed as shown in Fig. 8(a). At the feature mode, large column FPN was visible as shown in Fig. 8(b).

IV. Summary and Future work

The CMOS image sensor which can generate both normal image and feature data is proposed to reduce the power consumption of the image recognition system and the sensor output data size. Simulation results with 3-layer CNN shows that the recognition accuracy of the feature data is 96.9% and the data size is reduced by 99.0%, in which the original INRIA person dataset was converted to 3 bit horizontal edge dataset. Since the feature data can be aggressively quantized, the pixel source follower and the SAR-ADC of the column signal chain process the difference of vertically adjacent pixel inactivating the bias current and the total current consumption is only 0.371 [μA] for 5 bit feature data.

A test chip of the variable SAR-ADC was implemented with 0.18 μm CMOS process and was evaluated. The effect of the column FPN to the classification with the CNN will be verified in future. A CMOS image sensor with the proposed signal chain will be fabricated so that layer and filter channel structure of the CNN will be studied to reduce the power consumption of the image classification system also in future.

V. Acknowledgments

References

[1] S. Yoneda, Y. Negoro, H. Kobayashi, K. Nei, T. Takeuchi, M. Oota, T. Kawata, T. Ikeda, and S. Yamazaki, "Image sensor capable of analog convolution for real-time image recognition system using crystalline oxide semiconductor fet," in *International Image Sensor Workshop (IISW)*, pp. 322–325, 2019.

[2] C. Young, A. Omid-Zohoor, P. Lajevardi, and B. Murmann, "A data-compressive 1.5/2.75-bit log-gradient qvga image sensor with multi-scale readout for always-on object detection," *IEEE Journal of Solid-State Circuits*, vol. 54, no. 11, pp. 2932–2946, 2019.

[3] "Inria person." https://paperswithcode.com/dataset/inria-person. Accessed: 6/4/2023.

[4] M. Guy, B. Jan, W. Xinyang, and G. Vanhorebeek, "Backside illuminated global shutter cmos image sensors," in *International Image Sensor Workshop (IISW)*, no. R51, 2011.

[5] T. Sugiki, S. Ohsawa, H. Miura, M. Sasaki, N. Nakamura, I. Inoue, M. Hoshino, Y. Tomizawa, and T. Arakawa, "A 60 mw 10 b cmos image sensor with column-to-column fpn reduction," in *2000 IEEE International Solid-State Circuits Conference. Digest of Technical Papers*, pp. 108–109, 2000.

[6] K. Itsuki, A. Otani, H. Ogawa, and S. Okura, "A variable-resolution sar adc with 10-bit image capturing mode and 5-bit feature extraction mode," in *5th International Workshop on Image Sensors and Imaging Systems (IWISS2022)*, 2022.
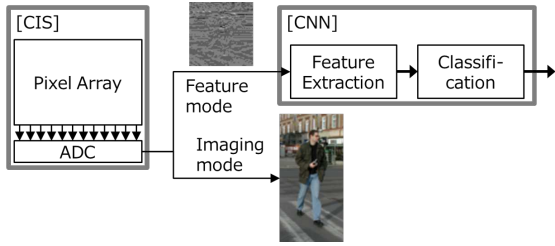
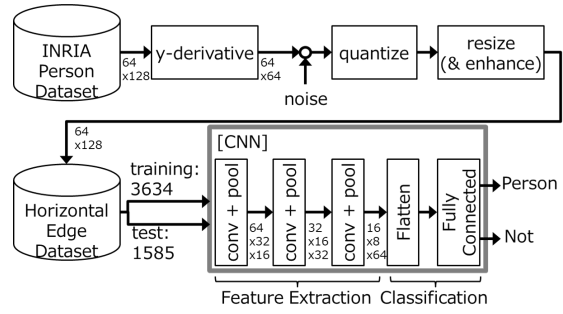Fig. 1. Overview of an image classification system with feature extractable CIS

Fig. 2. Block diagram of feature dataset conversion and a 3-layer CNN

	(1)	(2)	(3)	(4)
training	RGB	RGB	Edge-1	Edge-2
test	RGB	Edge-1	Edge-1	Edge-2
accuracy	98.3%	47.2%	95.7%	97.0%

TABLE I
SIMULATION RESULTS OF IMAGE CLASSIFICATION.
EDGE-1: 8 BIT HORIZONTAL W/O CONTRAST ENHANCEMENT.
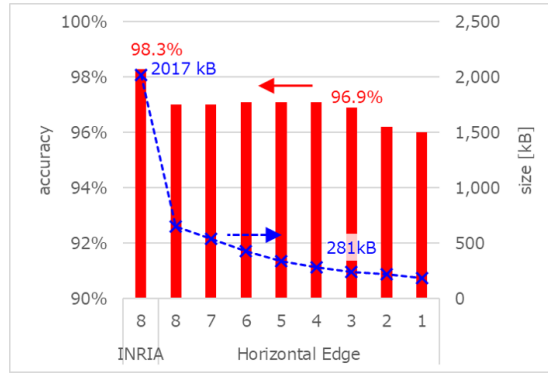EDGE-2: 8 BIT HORIZONTAL WITH CONTRAST ENHANCEMENT.

Fig. 3. Simulation results of image classification accuracy and data size
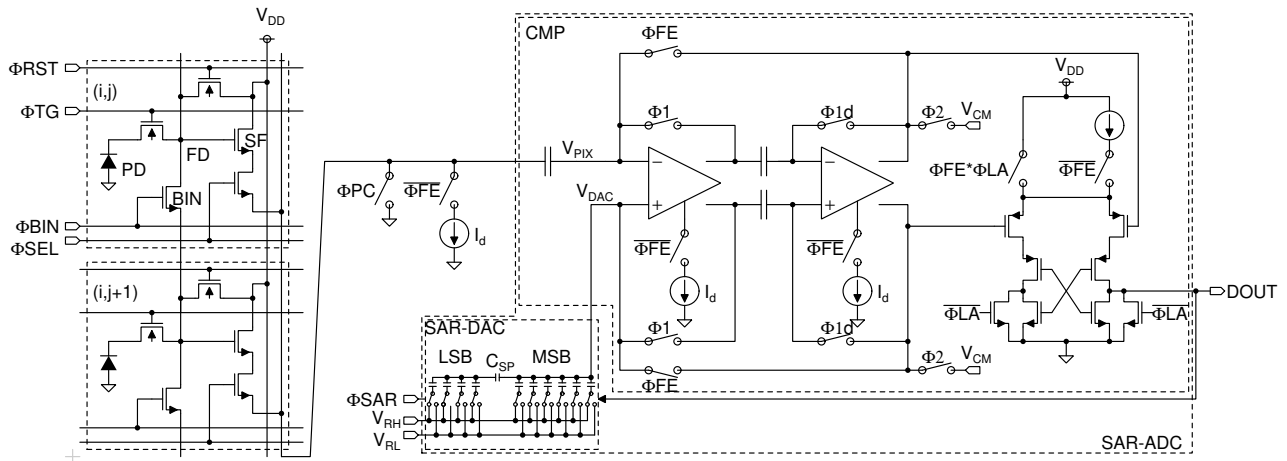

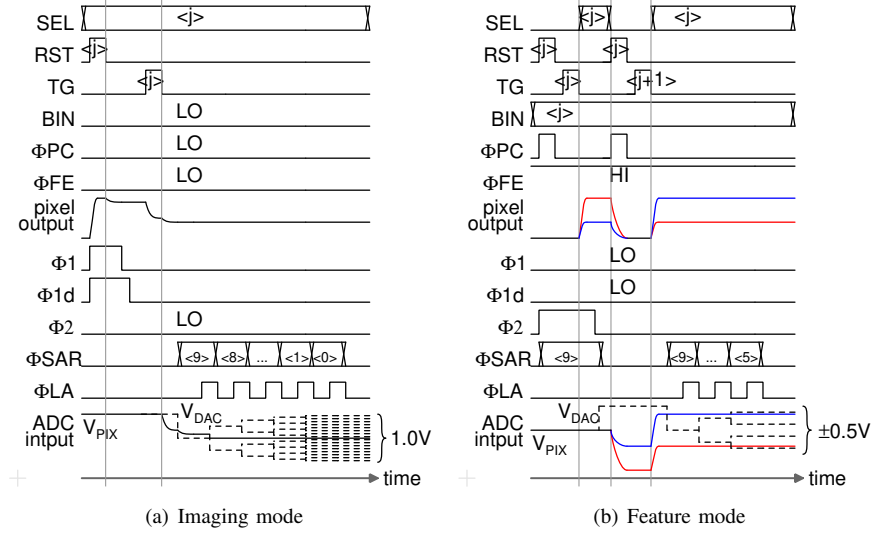Fig. 4. Schematic diagram of a feature extractable pixel and a variable SAR-ADC

(a) Imaging mode   (b) Feature mode

Fig. 5. Timing diagram

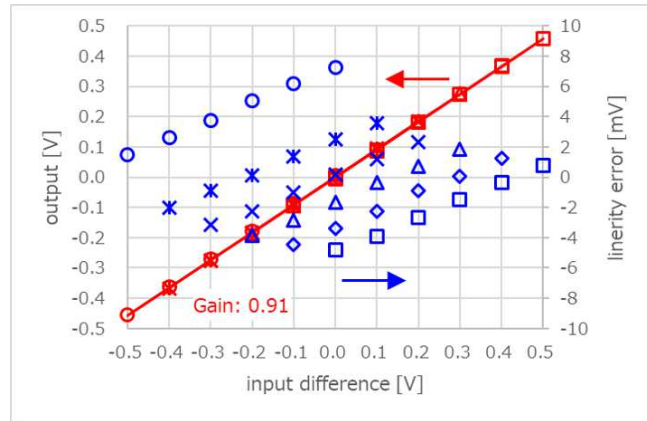| I_{dis} [μA]	Imaging mode	Feature mode
Pixel SF	9.2	0.26
ADC	24.1	0.057
total	33.3	0.317

TABLE II
SIMULATION RESULTS OF THE CURRENT CONSUMPTION

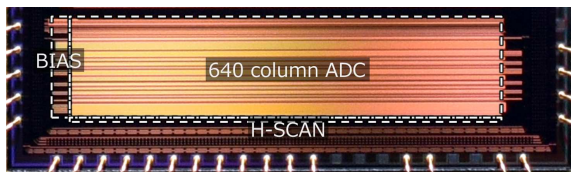Fig. 6. Simulation result of the pixel SF at the feature mode

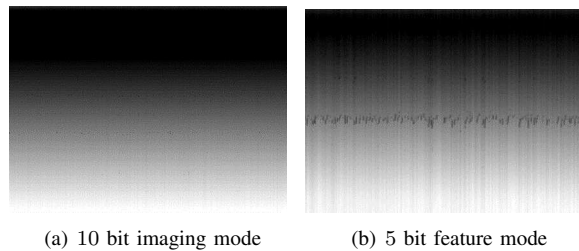Fig. 7. Chip photograph of the proposed variable SAR-ADC

(a) 10 bit imaging mode   (b) 5 bit feature mode

Fig. 8. Sample image for a ramp input signal

Temporal Noise Suppression Method using Noise-Bandwidth Limitation for Pixel-Level Single-Slope ADC

Sanggwon Lee, Min-Woong Seo, Masamichi Ito, Sung-Jae Byun, Hyukbin Kwon, Daehee Bae, Joosung Moon, Gihwan Cho, Heesung Shim, Jae-Kyu Lee, Chang-Rok Moon, and Hyoung-Sub Kim

Semiconductor R&D Center, Samsung Electronics Co. Hwasung-City, Gyunggi-do, 18848, Republic of Korea.

*Corresponding Author: minwoong.seo@samsung.com

Abstract— **This paper proposes a method for reducing temporal random noise (RN) in a pixel-level single-slope (SS) analog-to-digital converter (ADC) of a global-shutter (GS) CMOS image sensor (CIS). Pixel-level SS ADC is the ultimate scheme to improve image performances such as image quality, speed, noise, dynamic range, and power consumption compared to the column-level ADC. A temporal RN of the readout circuit represents flicker noise and thermal noise in the frequency domain. The noise-bandwidth limiting (nBWL) technique in the SS ADC circuit is a well-known method to limit the high-frequency response. By decreasing the -3dB cutoff frequency of the operational transconductance amplifier (OTA) for the comparator, the effective noise bandwidth is changed, and as a result a large portion of the thermal noise is filtered out. In order to achieve a low temporal noise global shutter imager, we have developed the pixel-wise ADC architecture with a 2-stacked structure including a large BWL capacitor. By finding an optimized size of the BWL capacitor to improve the RN performance, a precise circuit simulation and the related measurement were implemented. Because of the nBWL effect, the RN of the developed GS CIS with the BWL capacitor, approximately 100fF, has been reduced over 30% compared with the GS CIS without BWL capacitor. This noise reduction method is applied to our digital pixel sensor (DPS), which has a 4.95-μm pixel pitch and 2-megapixel (Mp) resolution, and the DPS has been successfully evaluated and demonstrated.**

Keywords—CMOS Image sensor, global shutter, pixel-level ADC, noise bandwidth limitation, temporal random noise, flicker noise, thermal noise, single-slope ADC

I. INTRODUCTION

Recently, the demands for the CMOS image sensor (CIS), especially a global shutter (GS) CIS, are increasing not only for mobile camera applications but also for machine vision image sensor applications such as an augmented reality (AR)/virtual reality (VR)/merged reality (MR), a security, and the automotive products. Until now, one of the main streams for CIS industry was a higher pixel resolution (larger than 200Mp) and a pixel shrinkage (less than 1.0 μm) based on the conventional column-parallel readout architecture [1-3]. In this case, however, some image degradation, such as dark shading, fixed-pattern noise (FPN), large power consumption, sensor noise, and so
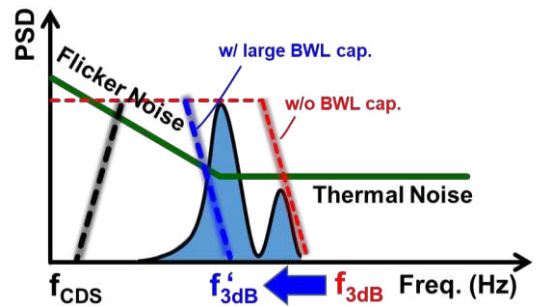

Figure 1. Conceptual noise power spectral density with/without nBWL technique.

on, are still issued, and to solve the problems, we need next-generation readout architecture like as pixel-parallel readout scheme [4-6]. A single-slope (SS) analog-to-digital converter (ADC)-type readout scheme, which is mainly utilized for the consumer imaging devices, is excellent candidate for the pixel-level readout approach. Because of limitation of pixel area, only a very simple amplifier and ADC configuration can be implemented. The SS ADC, which compares the integration of the input with a reference level and measures the time the integrator takes to reach this reference level, is one of the simplest forms of integrating ADCs. In spite of the area advantage of SS ADC for pixel-level readout scheme, the ADC noise is basically degraded by the size shrinkage of the comparator, normally 20 times smaller than that of the conventional. To minimize ADC noise, the noise-bandwidth limiting (nBWL) technique is utilized to the developed digital pixel sensor (DPS) [7-8]. In general, nBWL method with column-parallel readout scheme used in a limited role to suppress the sensor noise including flicker and thermal noises, because of keeping the sensor operation speed. On the other hand, nBWL method with pixel-parallel readout scheme can be aggressively used for eliminating the noise components in high-frequency thermal noise region, because a longer A/D conversion time is secured by using global A/D conversion operation, not conventional row-by-row A/D conversion.
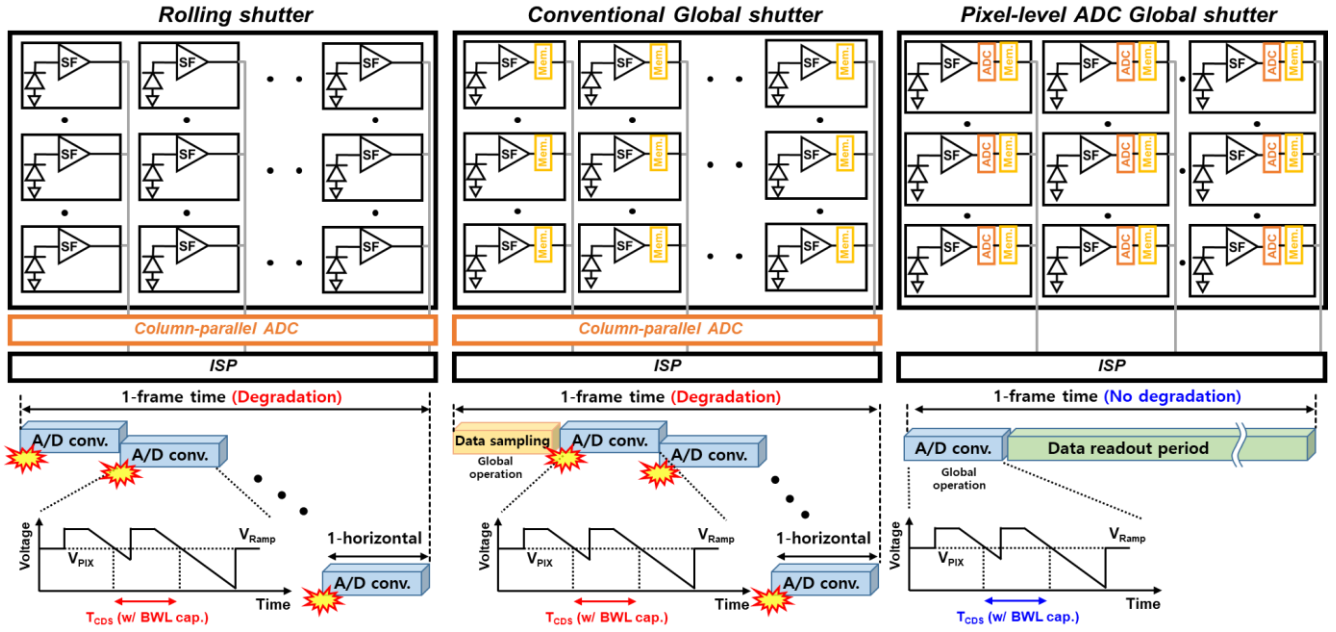
Figure 2. Sensor block diagrams with each simplified timing diagram for (a) Rolling shutter, (b) Conventional global shutter and (c) Pixel-level ADC global shutter architecture.

This paper has been verified using the transient-based AC simulation (TBAS) method [9] and demonstrated with the measurement results. The remainder of this paper is organized as follows. Section II describes the principle and architecture of the pixel-level SS ADC. Section III shows the simulation and measurement results with nBWL technique. Finally, the conclusion is given in Section IV.

II. OPERATIONAL PRINCIPLE AND ARCHITECTURE

A conceptual noise power spectral density (PSD) with/without BWL capacitor is shown in Fig. 1. In general, the temporal random noise (RN) of readout circuitry consists of two main components which are a flicker noise and a thermal noise in the frequency domain. Flicker noise is a low-frequency noise that has spectral density inversely proportional to the frequency. Flicker noise is caused by various sources such as amplifier's offset, kTC noise, and random fluctuations in the signals. By the correlated double sampling (CDS) function, the flicker noise in low-frequency is reduced and cancelled out. Thermal noise is mainly generated by the thermal motion of electrons in the transistor and it normally increase proportional to the circuit's bandwidth. Thus, a readout noise is finally decided by the both noise filters, the CDS function and -3dB cutoff frequency of comparator for SS ADC, as shown in Fig. 1. This paper describes the suppression of RN through the use of the BWL capacitor to shift lower band side for the -3dB cutoff frequency in the pixel-level ADC.
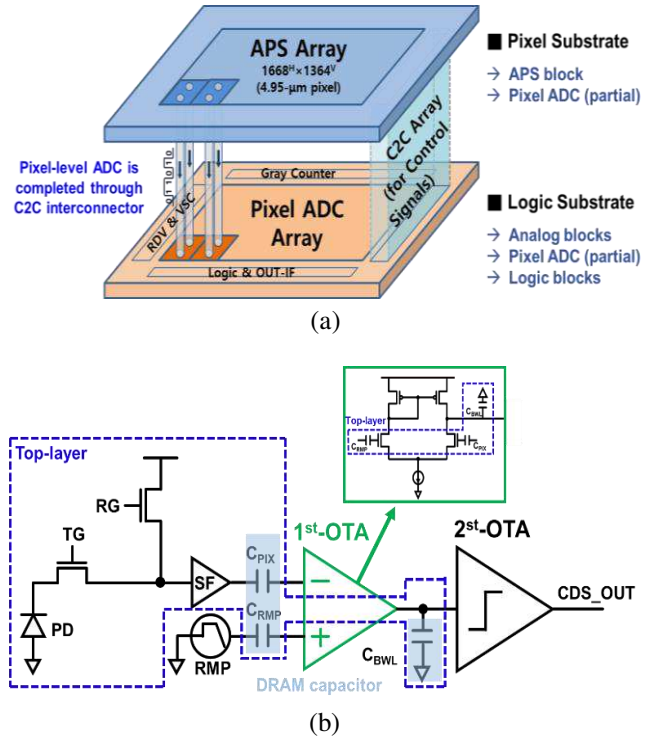

(a)

(b)

Figure 3. (a) Block diagram of the 2-stack structure and (b) Schematic of the pixel-level SS ADC.

The readout architectures of active pixel CMOS imagers are classified into three types as shown in Fig. 2, in largely where ADCs are located at columns and pixels. In Fig. 2(a), the readout architecture in rolling shutter, is commonly used in CMOS image sensors, performs A/D conversion start from the bottom of the sensor and top on the column ADC. Each row of pixels is exposed and readout sequentially, that creating a distortion effect on moving objects. The converted digital data is then sent out to the image signal processing (ISP). Fig. 2(b) shows
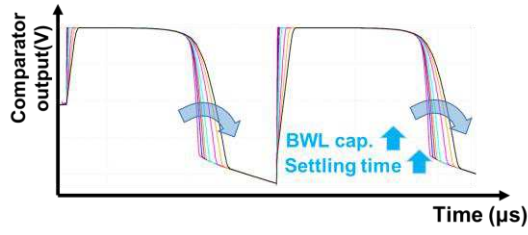
(a)
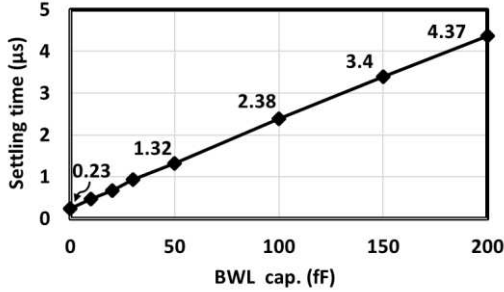
(b)

Figure 4. (a) Comparator output responses and (b) Signal settling time according to the capacitor's size of BWL.
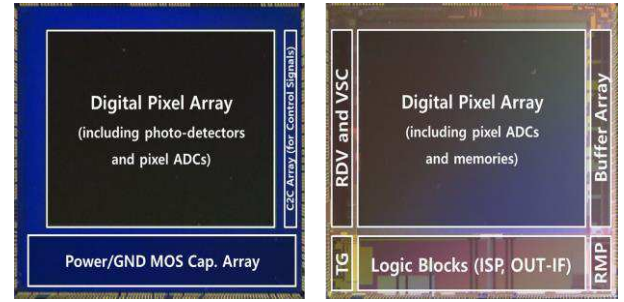

Figure 5. Chip micrographs (left: top chip, right: bottom chip).

Figure 6. The simulated temporal noises of pixel-level SS ADC with the different BWL capacitors.

the readout architecture in conventional global shutter. The exposure information, unlike rolling shutter CIS, is the same for 2-dimensional (2D) pixels and stored in-pixel memory with charge or voltage domain. This method eliminates the distortion effect in the rolling shutter. However, the readout method is exactly the same as rolling shutter CIS. In these method, the column ADC always operates during entire sensor operation. Therefore, the data conversion time is a key factor for the sensor speed in the column ADC architecture. On the other hand, the pixel-level ADC readout architecture performs analog-to-digital conversion on 2-dimensional pixel data in simultaneously as shown in Fig. 2(c). Then digitalized data is directly transfer ISP block. As a result, the ADC can have a sufficient time margin for the signal settle, and this architecture also can reduce the power consumption of each data converter. For the pixel-level SS ADC, even if it has the long term of A/D conversion period, sensor has no performance degradations such as a frame rate and noise. This means that the more aggressive nBWL technique can be used to the pixel-level readout circuitry for the noise suppression.

Fig. 3(a) and (b) show the block diagram of the developed 2-stack DPS structure and schematic of the pixel-level SS ADC, respectively. The prototype DPS is formed a 2-stack structure including photo-detector layer and analog/digital circuit layer. Top layer is composed of an active pixel sensor (APS) array with photo-diode and a part of the pixel-level SS ADC. Bottom layer consists of the rest of the pixel-level SS ADC array including in-pixel memory and the analog/digital circuits with a row driver (RDV) for the pixel signal control block, a vertical scanner (VSC) for choosing the pixel address number, a gray counter (GC),
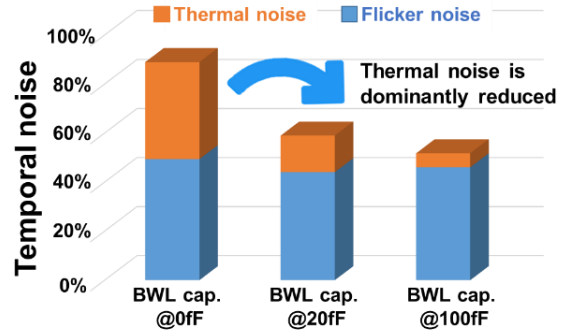
and a logic blocks for the image processing. Two layers including pixel-level ADCs have been bonded using Cu-to-Cu (C2C) process as shown in Fig. 3(a). Fig. 3(b) shows the simplified schematic of the pixel-level SS ADC. The input transistors of first-stage operational transconductance amplifier (OTA), the DRAM capacitors for AZ operation, and the DRAM capacitor for readout noise suppression using nBWL technique are implemented on the Top layer. A nBWL technique is one of the well-known noise suppression method for the readout circuit. However, depends on the size of BWL capacitor, the output settling time is gradually increased by increasing the output impedance. Fig. 4(a) and (b) show the output responses of comparator with different capacitor sizes and their settling times, respectively. At the capacitor size of 100fF, the total settling time of comparator output is expected around 4.8μs due to the reset and signal A/D conversions. A large settling time can limit the operation of image sensor.

III. SIMULTION AND MEASUREMENT RESULTS

A 2-Mp, 4.95-μm pixel pitch global shutter CMOS image sensor with pixel-level ADCs is implemented by using 65 nm (Top-layer with pixel part) and 28 nm (Bot-layer with logic part) CIS processes. The die micrographs of the top and bottom chips are shown in Fig. 5. The designed SS ADC has been verified using TBAS analysis method, and the sensor chip has been fully evaluated. The noise simulation results for demonstrating the
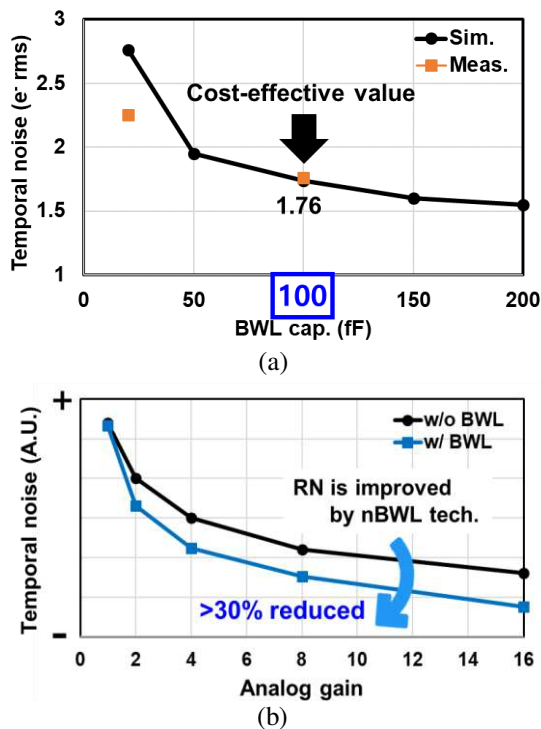
(a)

(b)

Figure 7. (a) The simulation and measurement results of sensor TN as a function of the capacitor's size for BWL (b) The measurement results of sensor TN as a function of the analog gain with and without BWL capacitor.

nBWL effect of pixel-level SS ADC are shown in Fig. 6. The thermal noise is sensitively decreased as a function of the BWL capacitor's size. As a result, a total RN is improved approximately over 30 % with the 100fF capacitor for nBWL. Fig. 7 shows the RN measurement results of fabricated DPS. The capacitor size of 100 fF is cost-effective value for pixel-level SS ADC with considering pixel pitch and sensor performances as can be seen from Fig. 7(a). Fig. 7(b) shows the measured noise trend with nBWL method as a function of analog gain. The measured RN of 1.76 e-rms is achieved with the 100fF BWL capacitor and the analog gain of 16 times. This result shows that the competitive low-noise pixel-level ADC can be realized, even if the comparator operates in sub-threshold region for the low-powered operation of the proposed readout architecture. The captured sample image is shown in Fig. 8. To analyze the RN trend with the nBWL technique, the top area of the image is assigned without BWL capacitor, and the right bottom and left bottom areas are assigned BWL capacitors of 100fF and 200fF, respectively. As can be seen from this sample image, the nBWL method is successfully worked and has been demonstrated without any image distortion.

IV. CONCLUSION

2-Mp GS-type with pixel-level ADC including RN reduction method has been presented and successfully

Figure 8. Captured sample image with BWL capacitor method.

demonstrated. We believe the pixel-level ADC will be a great alternative readout circuit solution to achieve the higher performance and functionality in the near future. The nBWL techniques with such readout chain are very effective method to improve the low-light performance of imager, especially digital pixel sensor (DPS) using pixel-level ADC.

References

[1] H. Kim et al., "5.6 A 1/2.65in 44Mpixel CMOS image sensor with 0.7μm pixels fabricated in advanced full-depth deep-trench isolation technology," in IEEE ISSCC Dig. Tech. Papers, pp. 1–3, Feb. 2020.

[2] Y. Nitta et al., "High-speed digital double sampling with analog CDS on column parallel ADC architecture for low-noise active pixel sensor," in IEEE ISSCC Dig. Tech. Papers, pp. 500–501, Feb. 2006.

[3] Martijn F. Snoeij et al.,"A CMOS imager with column-level ADC using dynamic column fixed-pattern noise reduction," IEEE J. solid state circuit, vol. 41, no. 12, pp. 3007-3015, Dec. 2006.

[4] M. Sakakibara et al., "A back-illuminated global-shutter CMOS image sensor with pixel-parallel 14b subthreshold ADC," in IEEE ISSCC Dig. Tech. Papers, pp. 79–81, Feb. 2018.

[5] T. Takahashi et al., "A stacked CMOS image sensor with array-parallel ADC architecture," IEEE J. Solid-State Circuits, vol. 53, no. 4, pp. 1061–1070, Apr. 2018.

[6] K. Mori et al., "A 4.0μm stacked digital pixel sensor operating in a dual quantization mode for high dynamic range," in Proc. Int. Image Sensor Workshop (IISW), pp. 308–311, Sep. 2021.

[7] M. Chu, et al., "An Extremely High-Speed and Low-Power Digital Pixel Sensor with Advanced Sensor Architecture," International Image Sensor Worshop (IISW), Sep. 2021.

[8] M.-W. Seo et al., "2.45 e-rms low-random-noise, 598.5 mW low-power, and 1.2 kfps high-speed 2-Mp global shutter CMOS image sensor with pixel-level ADC and memory," IEEE J. Solid-State Circuits, vol. 57, no. 4, pp. 1125–1137, Apr. 2022.

[9] H. Y. Jung et al., "Design and analysis on low-power and low-noise single slope ADC for digital pixel sensors," in Proc. Electron. Imag. (EI), Jan. 2022.

The source-to-gate capacitance of the in-pixel source follower: a positive feedback during charge sensing which increases column settling time and noise voltage.

Peter Centen, PeerImaging consulting, Mortelpleintje 1, 5051BW, Goirle, The Netherlands.
peter@peerimaging.com, +31 613679702.

Abstract

The gate-to-source capacitance of the in-pixel source follower, forms a capacitive attenuator with the floating diffusion capacitance. Voltage changes at the source are feedback to the gate: a positive feedback. It is effective when the resetfet is turned off. This is during charge sensing and reset level sensing for CDS. The positive feedback increases the output impedance at the source of the source follower and the noise voltage increases just as the settling time at the column.

Pre-amble: Simplifications

To focus on the main topic many simplifications have been applied. E.g. select and reset transistor are switches, simple MOS-transistor model for SF, no gate current noise, no bulk effect, noise excess factor of 2/3. Stationarity of the noise. The floating diffusion capacitance as a lumped form of the detection node capacitance. Column capacitance (pF) much larger than floating diffusion or SF gate-to-source capacitance (fF), one dominant pole. CDS applied. The intended approach is to keep it simple within reasonable restrictions that always apply when one has an imager that works.

Feedbackfactor

During charge sensing, the resetfet is off and the gate of the in-pixel source follower (SF), Figure 1, is floating, Figure 2 with RST=low. The gate-to-source capacitance (Cgs) has now an effective positive feedback to the detection node (Cfd), Figure 3. Consequently, any changes in the source voltage also changes the gate voltage. The amount of feedback from source to detection node depends on the ratio between gate-to-source capacitance Cgs and the floating diffusion capacitance Cfd,

$$\beta := \frac{Cgs}{Cfd}$$

During pixelreset (RST high, Figure 1,2) the gate of the source follower SF, and the capacitances connected to it, are clamped to a reference voltage, VDD-PIX. The gate of the source follower SF sees a low impedance. The capacitance between source and gate is a load for the source of SF and there is no feedback.

In an early paper on noise optimization [1] an optimal value for the gate-to-source capacitance was found Cgs=Cfd, or β=1. In a recent paper [2] a SF in a 45nm process with Cfd=1.34fF and Cgs=1.9fF had a slightly larger value of β=1.39. In general values of β=1/3 to β=2 will be found depending on the type of optimization [3,4,5] employed, eg 1/f, thermal, rtn.

Outputimpedance

In figure 4 the noise small signal equivalent diagram's are given for the reset state and charge sensing state. During the charge sensing state and seen from the SF source the positive feedback increases the output impedance with (1+β), Figure 2, 4

$$\frac{1+\beta}{gm}$$

In which gm is the SF transconductance.

Timeconstant

Figure 4 rightpart, with β=1 the RC-timeconstant for charging the column capacitor (Ccol) is doubled (1+ β=2), an important aspect in highspeed applications and for settling behavior

$$\tau := \frac{Ccol}{gm} \cdot (1+\beta)$$

Noisebandwidth

For a single pole network the noise bandwidth Bn [6] relates to the timeconstant. Due to the positive feedback the timeconstant is increased with (1+β) compared to the resetfet-on state

$$Bn := \frac{1}{4 \cdot \tau} = \frac{1}{4} \cdot \frac{gm}{1+\beta} \cdot \frac{1}{Ccol}$$

Noise

Assuming that the resetnoise is suppressed with e.g. CDS, the noise at the column is mainly caused by the source follower (in_sf) and the current source CS (in_cs), Figure 3,4. Due to the positive feedback the noise voltage at the SF source increases to:

$$en^2 = \left(in\_sf^2 + in\_cs^2\right) \cdot \left(\frac{1+\beta}{gm}\right)^2$$

with β=1 the noise spectral density increases 4 times. Irrespective if the noise is thermal, 1/f, RTN etc.

Thermal noise

The SF thermal noise current density is: $in\_sf^2 = 4kT \cdot \gamma \cdot gm$ and for the current source: $in\_cs^2 = 4kT \cdot \gamma \cdot gcs$ where γ is the noise excess factor, γ=2/3 will be used even though larger values can be found [4,7]. It equates to:

$$en^2 = 4 \cdot kT \cdot \frac{2}{3} \cdot (gm + gcs) \cdot \left(\frac{1+\beta}{gm}\right)^2 = 4 \cdot kT \cdot \frac{2}{3} \cdot \left(1 + \frac{gcs}{gm}\right) \cdot \frac{1}{gm} \cdot (1+\beta)^2$$

with respective transconductances for the source follower (SF) gm and for the current source (CS) gcs. The noise spectral density has a multiplier $(1+\beta)^2$ and the noise bandwidth Bn has a devisor (1+β). The noise variance (σ^2) at the column is then

$$\sigma^2 = Bn \cdot en^2 = \frac{kT}{Ccol} \cdot \frac{2}{3} \cdot \left(1 + \frac{gcs}{gm}\right) \cdot (1+\beta)$$

Even though σ^2 is the noise power at the column, only at the first sample and hold after the column it will show up and will have this noise variance in each sample! Practically the transconductance of the current source is smaller than that of the SF and the product of 2/3*(1+gcs/gm) will be close to 1. Using the optimal value for Cgs=Cfd, β =1, the noise variance simplifies into

$$\sigma^2 = 2 \frac{kT}{Ccol}$$

A bit larger than the often-expected kT/Ccol, or after CDS 4*kT/Ccol.

Conclusion

During charge sensing, the reset-off state, settling is (Cgs/Cfd+1) ~ 2 times slower than what one would intuitively expect when using only the SF-transconductance and the column capacitor value. The same holds for the noise voltage at the column. The sampled thermal noise at the column is not the kT/Ccol but about (Cgs/Cfd+1)*kT/Ccol.

References

[1] Hynecek, IEEE Trans. Electron Devices, vol. ED-31, no. 12, pp. 1713-1719, Dec. 1984.
[2] Chao et. al, IISW2017.
[3] Fasoli et. al., IEEE Trans. Electron Devices, vol. ED-43, no. 7, 1073-1076, July 1996
[4] Boukhayma, ISBN 978-3-319-68774-5, Januari 2018
[5] Centen, IEEE trans. Electron Devices. Vol. ED-38, No. 5, pp. 1206-1216, May 1991
[6] Carlson, Communication Systems, ISBN 0-07-009957
[7] C. Enz, E. Vittoz, "Charge-Based MOS Transistor Modeling: The EKV Model for Low-Power and RF IC Design", Wiley, Hoboken, 2006, ISBN:9780470855416
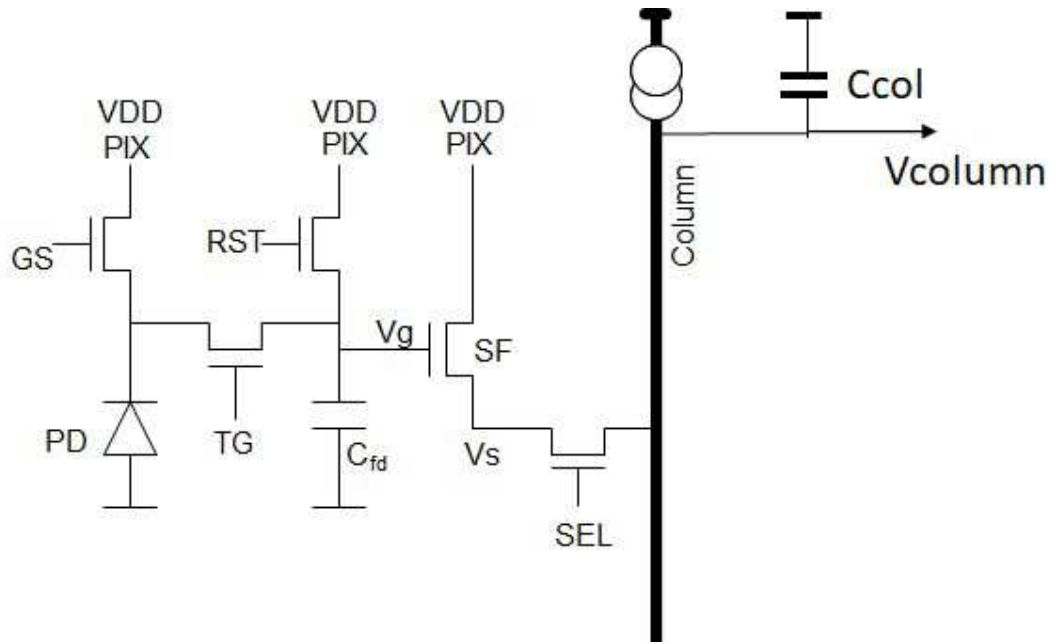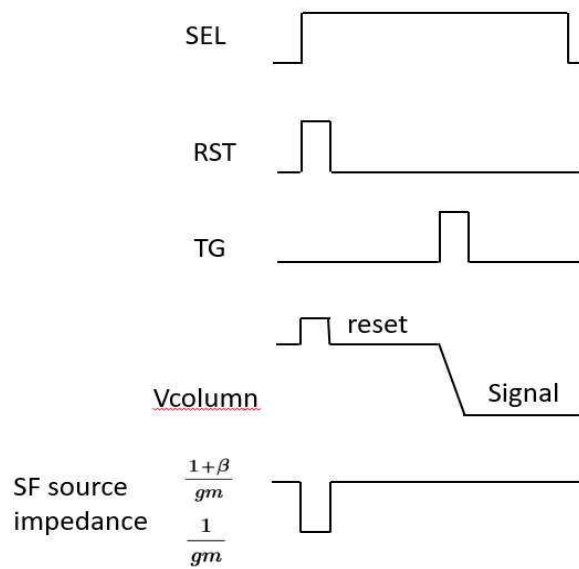
Figure 1 : 5T pixel

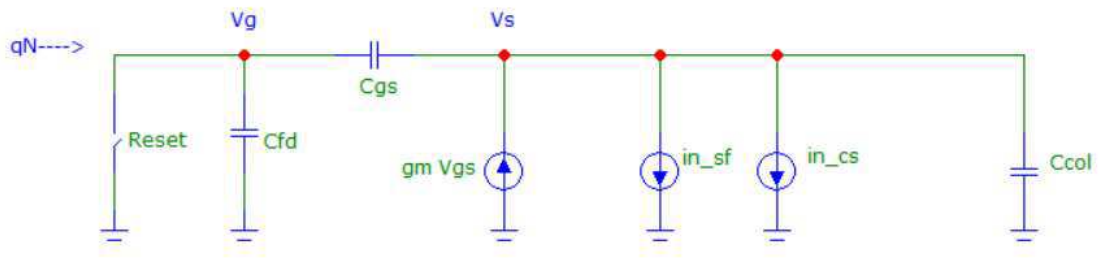*Figure 2: SF source impedance after reset and during charge sensing*

Figure 3: small signal diagram during charge sensing, reset-off

Figure 4: Noise small signal diagram. Left: reset-on, Right: reset-off

A Charge pump based TDI accumulator for CMOS Image Sensors

Rahul Kumar Singh[1,2,3], Siddhant Jain[2], Aakash Vishwakarma[2], and Mukul Sarkar[1,2]

[1]*Indian Institute of Technology, Delhi*
[2]*3rdiTech (DV2JS Innovation LLP), Delhi*
[3]*Email: rahul@3rditech.in, Tel: +91-1126591072*

Abstract—**Time delay integration (TDI) imaging sensors are used in remote push-broom sensing systems to improve the image quality in low light or when the relative speed of the scene and the detector is large. As the light from the scene accumulates in succession on each row, an accumulator adds the signals. The accumulator limits the signal-to-noise ratio of the TDI imaging systems. This paper presents an 8-stage, charge pump-based TDI accumulator addressing the saturation problem of integration-based accumulators. A prototype microchip of 128 x 8 TDI stage has been designed and fabricated in AMS 350 nm 1P4M OPTO process. The supply voltage used is 3.3 V with a pixel pitch of 10 μm. The measured SNR improvement for the 8-stages is 10.6 dB.**

Keywords— Time Delay Integration(TDI), CMOS Image Sensors(CIS), Analog accumulators

I. Introduction

Time delay integration (TDI) imaging sensors are used in remote push-broom sensing systems. It is used to improve the image quality in low light or when the relative speed of the scene and the detector is large. A large relative motion between the scene and the detector results in blurred images. In TDI, the pixels in the along-track direction capture the same target multiple times, therefore extending its equivalent integration time [1-2].

A typical 4-stage TDI operation has been demonstrated in figure 1. An object O1 is captured by pixels P1-P4 at time intervals $t_0 - t_3$. As the light from the scene is accumulated in succession on each row, an accumulator is used to add the signals. The extended integration time or accumulation of the same signal significantly improves the Signal-to-Noise (SNR) ratio of the captured image [3-4]. In the analog-domain accumulation, the output of the pixels are accumulated by an analog accumulator and then quantized by column ADC. The preferred choice of accumulator in the analog domain is switched capacitor-based integrator [2,3,5]. The maximum accumulated voltage for switched capacitor integrator-based accumulator depends on the supply voltage and thus has a limited accumulation and dynamic range.
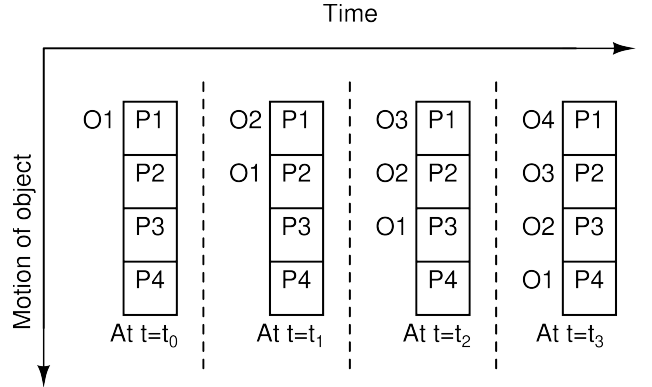

Figure 1: TDI operation

The multiple captures of the same scene work well in low-light environments or with a lesser number of TDI stages. In the presence of moderate or high light or as the number of TDI stages increases, the output of the accumulator saturates [2,5]. High precision and high linearity accumulation in low light have always been a problem with TDI [6]. Conventional integrator based accumulators have limited voltage swing and thus it becomes problematic in case of higher number of TDI stages or in case of over exposure of signals.

In this paper, a charge pump based hybrid TDI accumulator is presented. Charge pump based accumulator helps in utilising a hybrid mode of accumulation where analog as well as digital summation can happen. The hybrid algorithm also solves the accumulator saturation problem often seen with switched capacitor-based integrator. The hybrid mode extends the cumulative range of the TDI accumulator improving the SNR. A prototype chip has been designed and characterized in AMS 350 nm 1P4M OPTO process. The SNR boost obtained using charge pump-based accumulator is 10.6 dB for an 8-stage TDI.

The rest of the paper is organized as follows: section II describes the imager architecture, section III describes the measurement results and conclusions are presented in section IV.
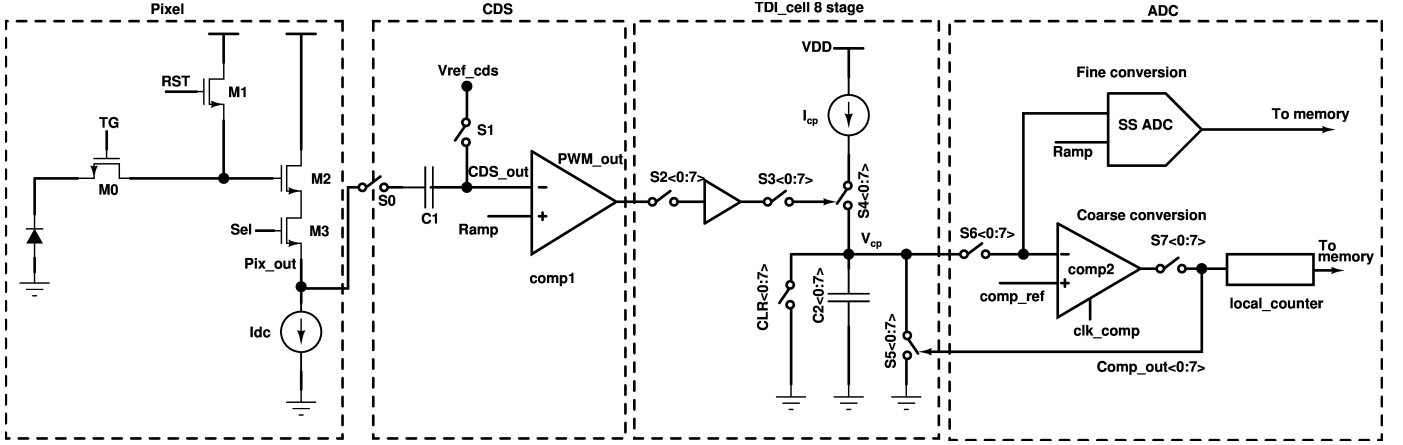
Figure 2: Proposed TDI architecture

II. System Architecture and circuit

The proposed system architecture for the linear TDI sensor is shown in figure 2. The TDI sensor consists of four components (a) 3T-APS (active pixel sensor) (b) correlated double sampling (CDS) with pulse width modulator (PWM) in column (c) charge pump-based integrator and (d) slope-based ADC for digital conversion. A 3T pixel with an extra switch between the photodiode and the source follower is used to decouple the sensing and storage nodes. The operational timing diagram for the sensor is shown in figure 3.

The integrated output of the pixel is transferred to column using in-pixel source follower. The output of the column is sampled on the CDS capacitor (C_1). The sampled output is compared with a ramp in comparator (comp1) to get a PWM signal. The width of the PWM signal represents the incident light intensity on the photodiode. The resultant PWM_out signal controls the charge pump. This charge pump functions as a TDI accumulator or TDI cell. When PWM_out signal is high, capacitor bank C_2 is charged with a DC current source I_{cp}. The accumulated output of the charge pump (V_{cp}) is compared with a fixed reference comp_ref in a clocked comparator (comp2).

A feedback circuitry using comparator comp2 is used to monitor the output of the charge pump (V_{cp}). It resets C_2 through switch S_5, when V_{cp} exceeds comp_ref value. The output of comp2 is stored in a 6 bit counter which tracks the total number of comparator triggering. The counter output is stored in a memory and provides the 6 MSB bits coarse conversion of the ADC. For PWM_out signal being active, the charge pump accumulates. When PWM_out is inactive, the charge pump holds its output V_{cp}. The stored V_{cp}, is quantized for fine conversion using a single slope 8-bit ADC. The resultant digitized signal is the combination of the 6-bit coarse and 8-bit fine conversion. The effective ADC resolution is 14 bits.
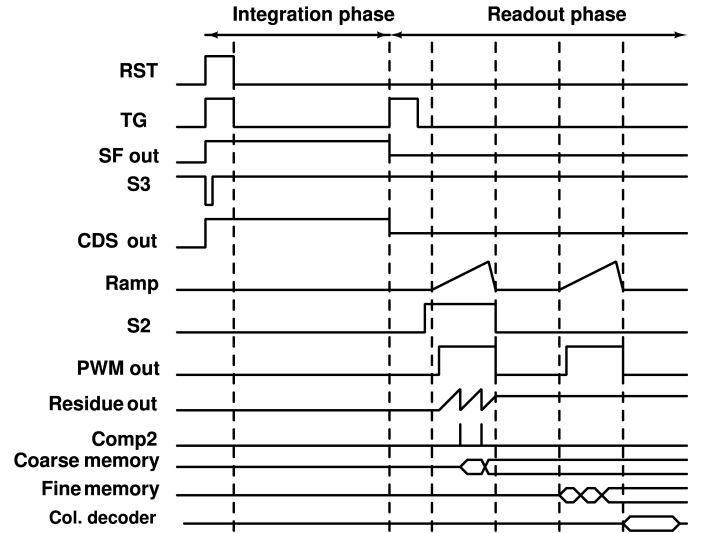

Figure 3: TDI operation

A. Charge pump accumulator theory

For high light, comp2 can trigger multiple times. It results in multiple integration cycles for C_2. Thus total pulse width of PWM signal is given as

$$T_{PWM} = nT_{ch} + T_{res} \qquad (1)$$

where n is the number of times capacitor C_2 is reset, T_{ch} is the charging time till V_{cp} reaches comp_ref value and T_{res} is the time left in PWM pulse after multiple reset signals.

Since the charge pump feeds the input to both analog and digital mode of accumulator conversion, the total digitized output is the sum of both analog as well as digital outputs. Thus digitized output is given as

$$(Vout)_{eq} = (Vout)_{ana} + (Vout)_{dig} \qquad (2)$$

where, $Vout_{ana}$ is the analog output which is the residue voltage stored on C_2. $Vout_{dig}$ is the voltage
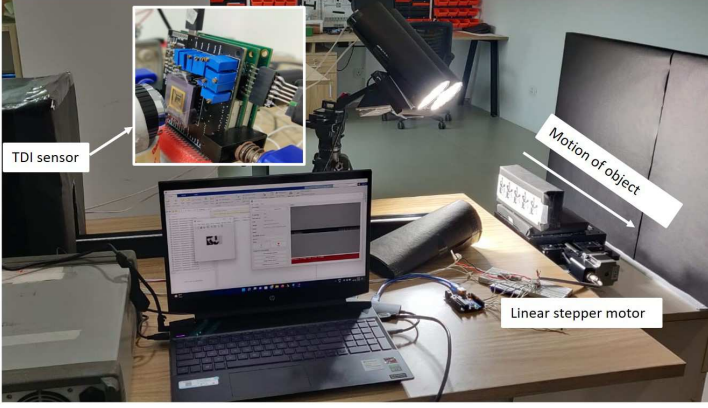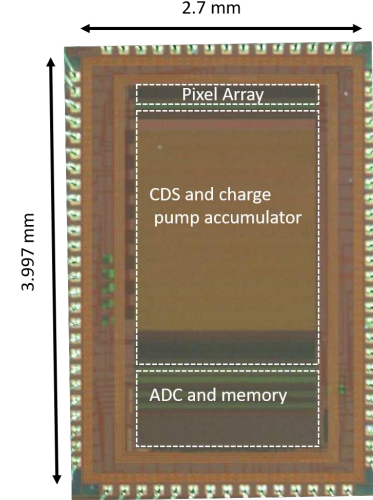
Figure 4: Measurement setup
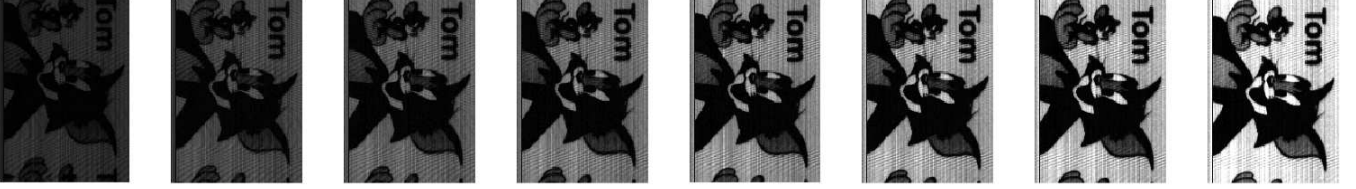

Figure 5: Microchip photograph of the sensor

Figure 6: Captured images using prototype sensor (TDI 1 to 8 stage)

output equivalent to the number of times the charge pump is reset. The analog output can be written as

$$(Vout)_{ana} = \int_0^{T_{res}} \frac{I_{cp}}{C_2}\, dt = \frac{I_{cp}T_{res}}{C_2} \qquad (3)$$

In case of coarse conversion digital equivalent voltage $Vout_{d1}$ can be given as

$$V_{d1} = \int_0^{T_{ch}} \frac{I_{cp}}{C_2}\, dt + \int_0^{T_{del}} \frac{I_{cp}}{C_2}\, dt \qquad (4)$$

where T_{ch} is the charging time till V_{ch} reaches comp_ref value, T_{del} is the excess delay time because of comparator delay. Excess delay term will come into picture only when residue voltage exceeds comp_ref and comparator gets triggered.

Thus digital equivalent of voltage that can be stored on C_2 when comparator triggers is

$$V_{d1} = \frac{I_{cp}(T_{ch} + T_{del})}{C_2} \qquad (5)$$

I_{cp} keeps on charging V_{cp} even when comp2 is high and V_{cp} is grounded. This results in a loss of charges which should have been part of the accumulation. Thus for multiple integration cycles, the equivalent digital output is given as

$$(Vout)_{dig} = n(V_{d1}) - \frac{(n-1)I_{cp}T_{clk}}{C_2} \qquad (6)$$

where T_{clk} is the comparator clock time when capacitor C_2 is forced to reset value while being charged

by the current source. The $n-1$ term in equation (6) denotes the loss of charges in accumulation.

$$(Vout)_{dig} = \frac{nI_{cp}(T_{ch} + T_{del})}{C_2} - \frac{(n-1)I_{cp}T_{clk}}{C_2} \qquad (7)$$

Thus the total digitized output can be written as

$$(Vout)_{eq} = \frac{I_{cp}T_{res}}{C_2} + \frac{nI_{cp}(T_{ch} + T_{del})}{C_2} - \frac{(n-1)I_{cp}T_{clk}}{C_2} \qquad (8)$$

Rearranging equation (8),

$$(Vout)_{eq} = \frac{I_{cp}}{C_2}(T_{res} + n(T_{ch} + T_{del}) - (n-1)T_{clk}) \qquad (9)$$

The comparator clock is chosen to be at a relatively higher frequency compared to the PWM. Thus $T_{clk} << T_{PWM}$. So equation (9) can be simplified as

$$(Vout)_{eq} = \frac{I_{cp}}{C_2}(T_{res} + n(T_{PWM} + T_{del})) \qquad (10)$$

The maximum output voltage, $(Vout)_{eq}$ is limited by a maximum swing in the analog domain but its digital equivalent is limited by the maximum count n that the counter can support. Thus total accumulation range can be enhanced n times by combining the coarse and fine conversion.
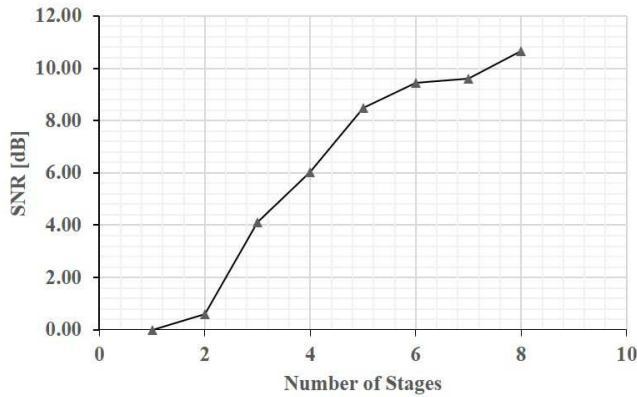
Figure 7: SNR boost plot

III. MEASUREMENTS RESULTS

Figure 4 shows the microchip photograph of the proposed TDI sensor. The TDI sensor is fabricated in AMS 350 nm OPTO process. It works on 3.3 V power supply. The pixel pitch is 10 μm x 10 μm. Figure 5 demonstrates the measurement setup. A linear actuator-based stepper motor has been used to provide orthogonal movement of the object with respect to the sensor. Corresponding images have been captured using Pleora iPORT CL-U3 frame grabber.

Figure 6 shows the images captured using the prototype sensor. These images are unprocessed images. As the TDI stage increases, the captured images have an improved SNR as expected. The accumulator suffers from non-idealities of the charge pump and hybrid ADC architecture. Therefore, the effective ADC resolution measured is 12-bit. The measured performance of the prototype sensor is summarized in Table. I. Figure 7 shows the measured SNR boost plot. The measured SNR with 8-stage TDI demonstrates an SNR improvement of 10.6 dB. The SNR boost obtained in [6] is 9.2 db for 8-stage TDI. Power dissipation per column is 16.5 μW at a line rate of 620 Hz.

Table I: Sensor characteristics

Technology (nm)	350
Pixel type (µm)	3T APS with extra switch
Array size	128 × 8
Pixel pitch (µm)	10
Supply voltage (V)	3.3
Max Line rate	1.07 kHz
Power consumption	16.5 µW per column @ line rate of 620 Hz
SNR boost	10.6 dB

IV. CONCLUSION

A prototype of 128×8 TDI sensor has been designed and fabricated in AMS 350 nm 1P4M OPTO process. The TDI imager uses charge pump based accumulator. The accumulator is reset based on the PWM signal generated proportional to the incident light intensity. The combination of the number of times the charge pump is reset and the residue voltage left on the charge pump gives a 12-bit resolution. The limitation in the resolution is due to the noise of the comparator. The proposed imager overcomes the swing limitation of conventional integrator based imagers by utilising hybrid mode of accumulation and provides a 10.6 dB SNR boost.

REFERENCES

[1] M. G. Farrier and R. H. Dyck, "A Large Area TDI Image Sensor for Low Light Level Imaging," in *IEEE Journal of Solid-State Circuits*, vol. 15, no. 4, pp. 753-758, Aug. 1980, doi: 10.1109/JSSC.1980.1051465.

[2] G. Lepage, J. Bogaerts and G. Meynants, "Time-Delay-Integration Architectures in CMOS Image Sensors," in *IEEE Transactions on Electron Devices*, vol. 56, no. 11, pp. 2524-2533, Nov. 2009, doi: 10.1109/TED.2009.2030648.

[3] H. Yu, X. Qian, M. Guo, S. Chen and K. S. Low, "A time delay integration CMOS image sensor with online deblurring algorithm," *VLSI Design, Automation and Test(VLSI-DAT)*, 2015, pp. 1-4, doi: 10.1109/VLSI-DAT.2015.7114510.

[4] K. -L. Liu, C. -C. Hsieh, S. -Y. Lai and C. -F. Chiu, "A time delay multiple integration linear CMOS image sensor for multispectral satellite telemetry," 2016 *IEEE Asian Solid-State Circuits Conference (A-SSCC)*, 2016, pp. 37-40, doi: 10.1109/ASSCC.2016.7844129.

[5] H. Yu, X. Qian, S. Chen and K. S. Low, "A Time-Delay-Integration CMOS image sensor with pipelined charge transfer architecture," 2012 *IEEE International Symposium on Circuits and Systems (ISCAS)*, Seoul, Korea (South), 2012, pp. 1624-1627, doi: 10.1109/ISCAS.2012.6271566.

[6] K. Nie, S. Yao, J. Xu, J. Gao and Y. Xia, "A 128-Stage Analog Accumulator for CMOS TDI Image Sensor," in *IEEE Transactions on Circuits and Systems I*: Regular Papers, vol. 61, no. 7, pp. 1952-1961, July 2014, doi: 10.1109/TCSI.2014.2304663.

Understanding 3D Imaging Performance in Sensors with Angle-Sensitive Pixels

Pascal Grégoire*, Niloufar Faghihi, Alexandre Favron, Gil Summy

Airy3D, Montréal (Qc) Canada
*Email: pascal.gregoire@airy3d.com

Abstract- **Image sensors equipped with angle-sensitive pixels (ASPs) can extract depth images at a fraction of the cost, resources, and power requirements of current solutions, but some technical challenges remain to obtain high accuracy 3D imaging. Off-the-shelf CMOS image sensors are upgraded into ASP cameras by the addition of a transmissive diffraction mask (TDM) patterned directly on top of the sensor. In this work, we model for the first time the depth sensitivity of ASP cameras. To demonstrate the validity of the model, the depth sensitivity is measured for diverse lenses and sensor architecture. This simple approach allows to maximise depth performances for a wide range of applications.**

Keywords—3D imaging, single-sensor, angle sensitive pixels, transmissive diffraction mask, PSF centroid

I. INTRODUCTION

Image sensors equipped with angle-sensitive pixels (ASPs) can extract depth information encoded into the defocus blur of typical 2D images. It is a simple approach compared to other types of 3D imaging, such as stereo vision or time-of-flight sensors, which rely upon several components (e.g., infrared emitters and receivers, multiple sensors). ASP cameras can be realized in several different ways. For example, Airy3D has developed a 3D imaging solution to generate near-field depth based on a transmissive diffraction mask (TDM) [1], [2]. The TDM is added above the microlenses of an image sensor (Fig. 1a) and modifies the angular response of each pixel to encode depth information. Recently, other approaches based on ASPs, such as dual-pixel (DP) cameras, have been

applied to depth map estimation [3]–[6]. Initially used for auto-focus [7], [8], DP requires careful design of microlenses, photodiodes and signal readouts. On the other hand, the addition of a TDM can transform an existing image sensor into a 3D sensor with no change to the pixel architecture. In both cases, the pixel angular response and the optical characteristics of the lens system determine the sensitivity of depth measurements. However, there is no clear description in the literature on how those parameters interplay [9], [10] and we are not aware of any predictive tool for depth performance.

In this work, we develop a simple and physically accurate model, based on a modified pillbox PSF, to quantify and optimize depth sensitivity. Section II describes the pixel structure and the link between 3D performance and the depth sensitivity, while Section III details the modified pillbox PSF model. Section IV shows the sensitivity measured on various 3D imagers, including a DP camera, and discusses the validity of the model using ray-tracing simulations.

II. PIXEL STRUCTURE AND 3D IMAGING

Fig. 1a shows a typical pixel structure for a TDM-based ASP camera. The TDM structure is added on top of an existing image sensor via standard mass-production techniques. It is composed of a few-microns thick spacer layer (pedestal) and a transmissive phase grating (TDM). Incoming light on the TDM experiences a phase modulation, which converts into an angle-dependent intensity modulation once the light propagates through the structure and is
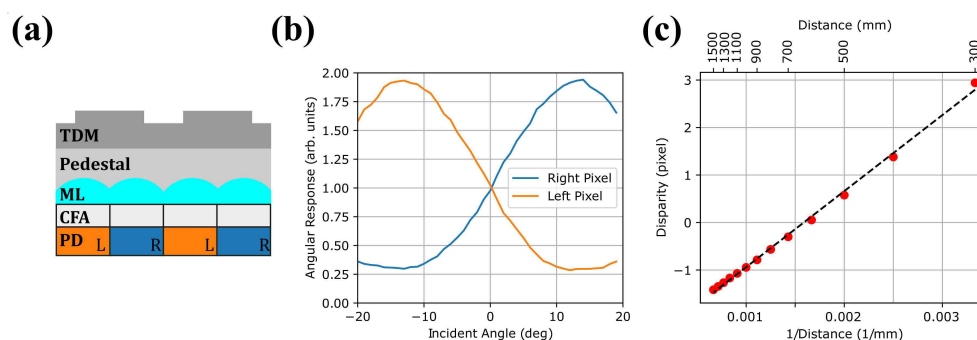

Fig. 1. (a) Schematic of an image sensor with a TDM. The photodiodes (PD), the color filter array (CFA) and the microlenses (ML) are unmodified components of an existing image sensor. A pedestal layer (light grey) and the transmissive diffraction mask (dark grey) are added above the microlenses. The TDM modifies the angular response of the photodiodes and leads to *left* (L) and *right* (R) pixels. **(b)** Measured angular response of left and right pixels for an image sensor with 1 μm pixel pitch. **(c)** Corresponding sensitivity curve using a 3.7mm F/2.0 lens. The measured disparity (red dots) is linear versus 1/distance. The slope of the linear fit (dashed line) is the depth sensitivity $S = 1601$ mm*pixel.
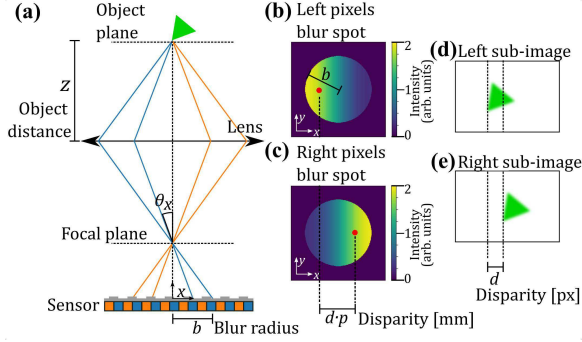
Fig. 2. (a) Thin lens model of an ASP camera. A point source at distance z from the lens forms a defocused spot of radius b at the sensor plane. The sensor is composed of alternating left (orange) and right (blue) pixels having angular responses as in Fig. 1b and a pixel pitch p. The blur spot intensity distribution of **(b)** the left pixels subset and **(c)** the right pixels subset are affected by the asymmetry of the angular response. A ray crossing the aperture on the right side of the lens reaches the sensor with a negative incident angle. The left blur spot is then more intense on the left side since the left pixels are more sensitive to negative angles. The red dots illustrate the PSF centroids, and their position difference is the disparity (Section III). Considering the image of an object placed at distance z, the left and right pixels form **(d)** a left and **(e)** a right sub-image, respectively, where the relative position of the object in the image is shifted by d pixels.

integrated by the photodiodes. The TDM effectively redistributes the light between adjacent pixels as a function of the incident angle, which leads to two subsets of pixels with asymmetric angular responses (Fig. 1b).

As illustrated in Fig. 2, each subset of pixels captures a different viewpoint of the scene, analogous to a stereo-camera system. It is then possible to measure the distance of an object based on its apparent displacement between the two sub-images (left versus right viewpoints, see Fig. 2d-e). This displacement, usually measured in pixels, is the disparity d and is related to the object distance z using

$$d = S(1/z - 1/z_F) \qquad (1)$$

where S is the depth sensitivity and z_F is the focus distance. As also shown in Fig. 1c, the disparity is inversely proportional to the distance and is zero when an object is in focus ($z = z_F$). Once the parameters S and z_F are determined, the measured disparity can be transformed into a depth measurement as shown in Fig. 3.

An error in the disparity evaluation Δd leads to a depth error Δz of the form

$$\Delta z = \frac{z^2}{S} \Delta d \qquad (2)$$

which shows that maximising the depth sensitivity S reduces the depth error (increases the precision). It should be noted that Δd depends on the type of algorithm to extract the disparity, amongst other factors, while the sensitivity S is entirely dependent on the camera system parameters (lens, pixel size, angular response of pixels).
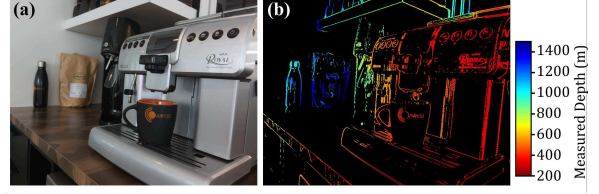

Fig. 3 Typical scene acquired with a TDM-enabled ASP camera using a 1.0 μm pixel Bayer sensor equipped with a $f = 3.7$mm F/2.0 lens. The color 2D image **(a)** and the depth map **(b)** are extracted from a single capture.

III. DEPTH SENSITIVITY MODEL

Having an accurate and simple model for the depth sensitivity S is then crucial to optimize a camera design and produce high quality depth maps.

The first step is to relate the point-spread function (PSF) of the camera to the disparity. Our starting point is the image formation model of a point source P, expressed as $I^{(i)} = P \otimes PSF^{(i)}$, where $i = L, R$ for the left or right pixels, $\otimes$ denotes a convolution and $I^{(i)}$ is the image of the point source for the left or right subset of pixels. One possible way to define the position of a defocused point source at the sensor plane is by its centroid. The disparity is then the displacement of the centroids between the left and right sub-images and expressed in pixel units p:

$$d = \frac{1}{p}\left(\langle I^{(L)}\rangle_x - \langle I^{(R)}\rangle_x\right) = \frac{1}{p}\left(\langle PSF^{(L)}\rangle_x - \langle PSF^{(R)}\rangle_x\right), \qquad (3)$$

the centroid along the x axis of a function $g(x, y)$ being defined as

$$\langle g \rangle_x = \frac{\iint x\, g(x,y)\, dx\, dy}{\iint g(x,y)\, dx\, dy} \qquad (4)$$

In (3), the second equality comes from the image formation model and the properties of convolutions [11]. This example can be generalized to any complex scene, not only point sources, meaning the centroid of the left and right PSFs are directly related to the disparity.

The next step is to define a model for the PSFs. We chose to use geometrical optics, for sake of simplicity, and adapted a pillbox PSF model [12]. First, assuming a thin lens approximation, there is a relationship between the angle of incidence θ_x of a ray and its position on the sensor plane x, which is related to the optical blur radius b:

$$b = \frac{f^2}{2N}(1/z - 1/z_F), \qquad f \ll z \qquad (5)$$

$$x = f^2(1/z - 1/z_F)\theta_x, \qquad \theta_x \ll 1 \qquad (6)$$

with f the focal length and N the f-number of the lens.

The model is simple, the PSF is approximated by a circle of radius b and its intensity is modulated by the
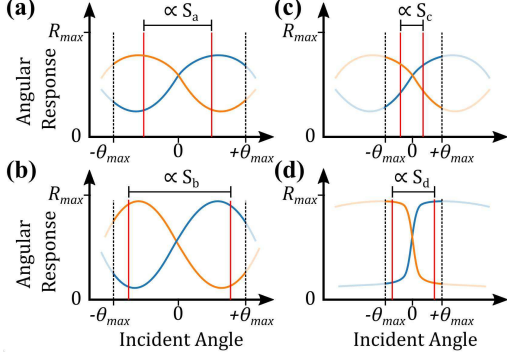
Fig. 4. Illustrating various typical PSF cross-sections in the angular space. The angular responses of the left pixels $R^{(L)}$ (orange) and the right pixels $R^{(R)}$ (blue) are shown. **(a)** The dashed lines are the maximal angles from the lens determined by the f-number and define the limits of the PSFs. The red lines illustrate the PSF centroid positions in the angle space $\langle \text{PSF}^{(L)}\rangle_{\theta_x}$ and $\langle \text{PSF}^{(R)}\rangle_{\theta_x}$, and their difference is proportional to the depth sensitivity S as in (9). **(b)** The amplitude of the angular response is bigger, giving a greater sensitivity. **(c)** A larger f-number reduces the sensitivity; the angular response is *clipped*. **(d)** The angular response shape maximizes the sensitivity for a larger f-number.

angular response of the pixel. The validity of this approximation is discussed in Section IV.

$$\text{PSF}^{(i)} = A(x, y, z) \cdot R^{(i)}\left(\theta_x, \theta_y\right) \tag{7}$$

where $R^{(i)}\left(\theta_x, \theta_y\right)$ is the angular response of pixels $i = L, R$ (potentially a 2D angular response), and where A is a step function defined by the blur radius b,

$$A(x, y, z) = \begin{cases} 1, & if \ x^2 + y^2 \leq b(z)^2 \\ 0, & elsewhere \end{cases} \tag{8}$$

with (x, y) the position on the sensor plane with respect to the optical axis. The left and right pillbox PSFs are illustrated in Fig. 2b-c, where (6) allows to express $R^{(i)}$ as a function of position instead of angle.

Combining (3) to (8), the disparity is

$$d = \underbrace{\frac{f^2}{p}\left(\langle \text{PSF}^{(L)}\rangle_{\theta_x} - \langle \text{PSF}^{(R)}\rangle_{\theta_x}\right)}_{S}(1/z - 1/z_F) \tag{9}$$

which has the same form as (1). Stated explicitly, the depth sensitivity is independent of the distance z

$$S = \frac{f^2}{p}\left[\langle a \cdot R^{(L)}\rangle_{\theta_x} - \langle a \cdot R^{(R)}\rangle_{\theta_x}\right] \tag{10}$$

with a the step function A now expressed with respect to angle (a is the numerical aperture):

$$a\left(\theta_x, \theta_y\right) = \begin{cases} 1, & if \ \theta_x^2 + \theta_y^2 \leq \theta_{max}^2, \\ 0, & elsewhere \end{cases} \tag{11}$$

with $\theta_{max} = 1/(2N)$. Equation 10 shows easily that a high focal length f and a small pixel pitch will maximise *S*. In practice, a long focal length is not always desirable since the focal length and the sensor size determine the angular field-of-view of the camera.

Balancing the sensor size (the cost), the field-of-view and the depth sensitivity is then highly application-specific.

The last term of (10) is illustrated in Fig. 4 and can be understood as follow: the angular response is only contributing within the numerical aperture (determined by the f-number N in our model), the rest outside $[-\theta_{max}, +\theta_{max}]$ is clipped by the lens. The centroid of this *clipped* angular response needs to be as off-axis as possible to maximize the difference between the left and right PSFs. It can be accomplished with a strong asymmetry in the angular response (Fig. 4b, high ratio between the minimum and the maximum of the responses) and/or by shaping the profile of the angular response curve. For example, Fig. 4d shows a sharp transition followed by a plateau up to the maximal angle, which is optimal to increase the centroids difference.

Figure 4 shows that increasing the f-number (reducing the aperture) generally means a smaller sensitivity. Notably, two regimes can be distinguished. When the angular response is approximately linear around $\theta_x \approx \pm\theta_{max}$, such as in Fig. 4c, the sensitivity will scale as $S \propto 1/N^2$. Instead, if the angular response is constant around $\theta_x \approx \pm\theta_{max}$, like in Fig. 4d, then $S \propto 1/N$.

In the general case, the angular response needs to match with the targeted f-number. Although the response in Fig. 4d is optimal for a wide range of f-numbers, it is not possible to obtain easily with most image sensors. A sinus-like response, as in Fig. 4a-c is the norm. There, if the peak of the angular response occurs outside the numerical aperture of the lens, the contribution from the peak is lost and the sensitivity is further reduced ($1/N^2$ regime) . It is possible to bring the peak of the angular response closer to $\theta_x = 0$ by adjusting the TDM design.

IV. RESULTS AND DISCUSSION

To illustrate the validity of our model, we measured the depth sensitivity of various ASP cameras; multiple TDM-enabled image sensors and a digital full-frame camera with a DP architecture. The TDMs were applied using standard CMOS processes onto four types of image sensors with distinct pixel architectures, spanning from a 3.2 μm pixel industrial sensor to a 1.0 μm pixel Bayer sensor for mobile applications. The cameras were focused at a distance $z_F = 500$ mm, and the depth sensitivity was measured by capturing a series of planar scenes at known distances z between 300 and 1500 mm. The disparity was extracted from each image using our custom algorithm (DepthIQ™). A linear fit of the disparity vs. $1/z$ curve gives the sensitivity S (see Fig. 1c). The sensitivity in the central portion of the field-of-view is reported in Fig. 5.

The modeled sensitivity needs the angular responses, which have been measured using the sensors without a lens. Each sensor is illuminated with a distant point
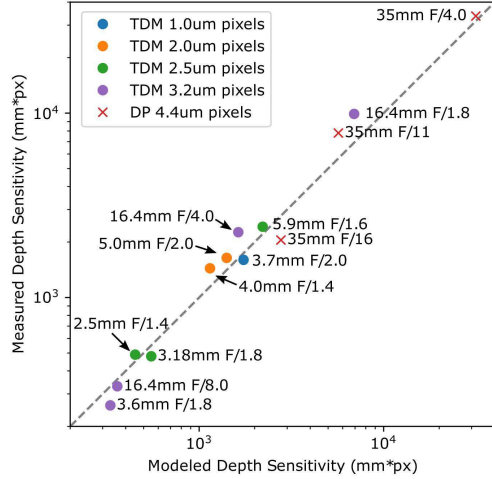
Fig. 5. Measured vs. modeled depth sensitivity for various sensor and lens combinations. The dashed line represents a perfect agreement between the measurement and the model. The measurement error is estimated to be ± 5%. Each TDM (dots) or DP (crosses) camera is labeled with the lens focal length and f-number. The depth sensitivity increases for higher focal lengths and lower f-numbers (larger apertures).

source, and the angle of the sensor plane is varied by 1 degree steps using a goniometer. The angular response is reconstructed from captures taken at each angle. As shown in Fig. 5, good agreement is found between the model and measured performances in all cases, over two orders of magnitude in depth sensitivity.

To confirm the validity of our adapted pillbox PSF model, we carried out a ray tracing simulation using Zemax OpticStudio. An F/1.8 lens with a 16.4 mm focal length (Edmund Optics 86-571) was combined with a custom layer that modulates the ray intensity at the sensor plane according to the angle of incidence. The PSF of the left and right pixels was simulated using the angular response measured from the 3.2 μm TDM-enabled sensor. Figure 6 compares the PSF of the left pixel to the case without ASPs (flat angular response). For the out-of-focus distance (Fig. 6b-c), the PSF shape is close to the geometrical optics regime. The impact of the TDM is almost exclusively a modulation of the intensity, which is similar to the pillbox model behavior. At the focus distance (Fig. 6d-e), geometrical optics is no longer valid, and the angular response is no longer visible; the PSFs with or without ASPs are identical. Even if the pillbox model is inexact in this regime, the disparity extracted by the PSF centroids as in (3) is still linear in 1/distance (Fig. 6a). In fact, the sensitivity predicted by the pillbox model (S=6920 mm*px) is very close to the sensitivity using the complete ray-tracing simulation (S=6770 mm*px). What matters is not the exact shape of the PSFs, only their centroid position, which is well captured by the pillbox model at all distances.

V. CONCLUSION

We have presented a complete model for the depth sensitivity of an ASP camera, based on the centroid of
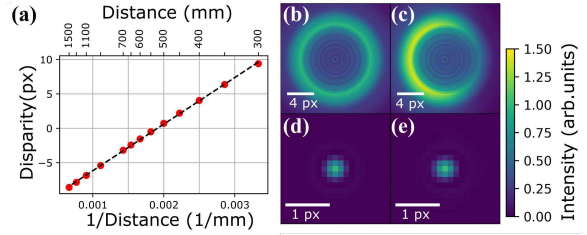

Fig. 6. Zemax OpticStudio simulation using an F/1.8 lens with a 16.4mm focal length and the 3.2 μm pixels TDM-enabled sensor. **(a)** Simulated disparity using the OpticStudio PSFs at various distances. The disparity is extracted using (3). The dashed line is a linear fit and its slope gives the depth sensitivity S=6770 mm*px. **(b)** and **(c)** are simulated out-of-focus PSFs (object distance z = 400mm, focus at z_F = 500mm) without and with ASPs, respectively. **(d)** and **(e)** are simulated PSFs at the focus distance z = z_F = 500mm without and with ASPs, respectively.

a modified pillbox PSF, and confirmed it accurately describes the 3D imaging capabilities of dual-pixel and TDM-based cameras. It highlights the trade-offs when optimizing the depth performance of an ASP camera for a specific application, notably the impact of the aperture size and the desired angular field-of-view onto depth sensitivity. Our approach is then a necessary tool for designing high performance angle-sensitive 3D imager.

REFERENCES

[1] N. Kunnath, "Depth from Defocus using Angle Sensitive Pixels based on a Transmissive Diffraction Mask," M.S. thesis, McGill University, Montreal, 2018.

[2] G. Summy and J. Mihaychuk, "Diffraction mask design brings 3D imaging to standard CMOS image sensors," *Laser Focus World*, 2020. Accessed: Dec. 05, 2022.

[3] R. Garg, N. Wadhwa, S. Ansari, and J. T. Barron, "Learning Single Camera Depth Estimation using Dual-Pixels," in *ICCV*, Apr. 2019.

[4] A. Punnappurath, A. Abuolaim, M. Afifi, and M. S. Brown, "Modeling Defocus-Disparity in Dual-Pixel Sensors," in *ICCP*, 2020.

[5] N. Wadhwa *et al.*, "Synthetic depth-of-field with a single-camera mobile phone," *ACM Trans Graph*, vol. 37, no. 4, 2018.

[6] S. Xin *et al.*, "Defocus Map Estimation and Deblurring from a Single Dual-Pixel Image," in *ICCV*, 2021. Accessed: Dec. 04, 2022.

[7] E. S. Shim *et al.*, "All-Directional Dual Pixel Auto Focus Technology in CMOS Image Sensors," in *IEEE Symposium on VLSI Circuits, Digest of Technical Papers*, Jun. 2021, vol. 2021-June.

[8] M. Kobayashi *et al.*, "A Low Noise and High Sensitivity Image Sensor with Imaging and Phase-Difference Detection AF in All Pixels," in *IISW*, 2015. Accessed: Dec. 05, 2022.

[9] B. S. Choi *et al.*, "Analysis of disparity information for depth extraction using CMOS image sensor with offset pixel aperture technique," *Sensors (Switzerland)*, vol. 19, no. 3, Feb. 2019.

[10] K. Fukuda, "A Compressed N×N Multi-Pixel Imaging and Cross Phase-Detection AF with N×1RGrB + 1×NGb Hetero Multi-Pixel Image Sensors," in *IISW*, 2021.

[11] E. W. Weisstein "Convolution." From MathWorld--A Wolfram Web Resource.
https://mathworld.wolfram.com/Convolution.html

[12] S. W. Smith "The scientist and engineer's guide to digital signal processing," San Diego: California Technical Publishing, 1997, pp.400-402.

A SPAD based Multi-Mode Compressive LiDAR Pixel for Depth Ranging by Direct ToF Measurement

Sohail Faizan, Kapil Jainwal - *Member IEEE*, Minal Bisen, and Nitin Khanna

Abstract—This work proposes a novel compressive Light De-tection and Ranging (LiDAR) pixel. The pixel operates by the direct Time-of-Flight (dToF) method and utilizes a Single Photon Avalanche Diode (SPAD) for photon detection. It implements a current-starved circuit for an initial logarithmic response, result-ing in a higher resolution at short ranges. The pixel subsequently switches to a linear circuit, providing a large full-scale range. The multi-mode implementation provides high accuracy and a long full-scale range (FSR), while avoiding a high data rate all in a compact pixel design. The proposed pixel is realized in standard 65nm TSMC process. It has a minimum detection distance of 16 mm (106.65 ps), a logarithmic depth resolution of 16mm - 0.17 m and a linear depth resolution 0.59 m (3.9 ns). The dynamic range of the pixel is 500 m. The maximum voltage error of the logarithmic part is 2.5% to -0.5%. The logarithmic pixel has a pitch of 8 μm×6 μm

I. INTRODUCTION

Depth measurement is a crucial mode of sensing, the need for which is rapidly growing with the ongoing automation revolution. The Time-of-Flight (ToF) method uses the time taken by LASER pulses reflected from a surface to calculate the distance. Indirect ToF (iToF) uses the phase difference between the transmitted and the received pulse to calculate the distance, while direct ToF (dToF) measures the total time taken by the light pulse from the transmission to detection [1]- [3]. iToF based pixels provide a higher resolution but are confined to small ranges, while dToF provides a lower resolution but can be designed for much larger ranges. The dToF ranging is realised using time-to-amplitude converters (TAC) or time-to-digital converters [4]- [15]. TACs integrate the time to a voltage, while TDCs use counters to measure the time [11]- [12]. The former are more compact but offer lower depth resolution, while the latter are larger (due to large in-pixel elements such as counters) but provide a high degree of depth accuracy [4]- [15].

Compressive pixels are used to achieve higher sensitivities at desired ranges [4]. An implementation of the same is logarithmic pixels (Fig. 1(a)), which provide high sensitivity at lower ranges. Standard logarithmic pixels are built by a current-starved design [4], where the gate of a transistor is connected to a full switch on voltage, but the transistor is kept in the sub-threshold region by limiting the current flowing through it. The sub-threshold operation gives a logarithmic voltage relation dependent on the current flowing through the circuit as follows:

$$V_{\text{out}} = V_{\text{dd}} - \alpha V_{\text{T}} \ln\left(\frac{L}{W} \cdot \frac{I_{\text{ds}}}{I_{\text{d0}}}\right), \qquad (1)$$
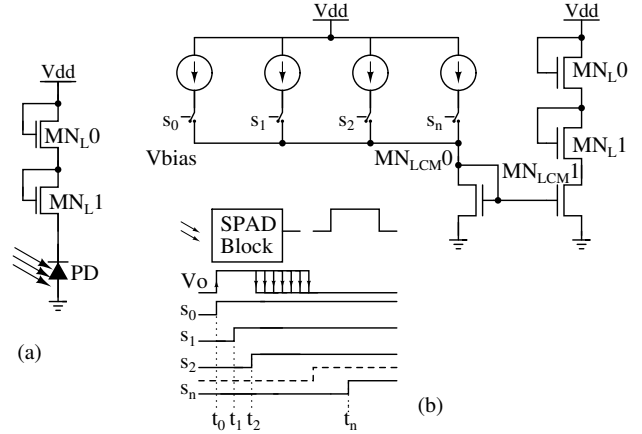

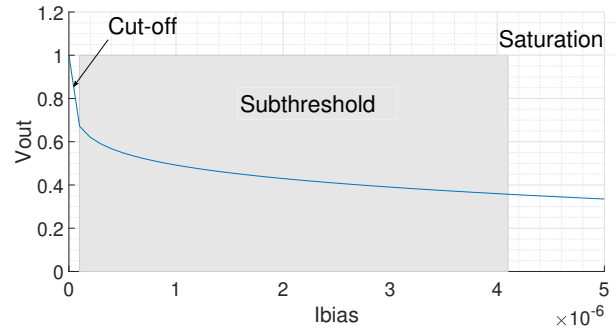Fig. 1: (a) Conventional log pixel (b) Step logarithmic gener-ation theorisation

Fig. 2: Operational range of log design

where V_{dd} is the supply voltage, V_{T} is the thermal voltage, L (length) and W (width) stand for the general representations of the dimensions of the transistor MN_L0, 'α' refers to the sub-threshold swing coefficient, and I_{d0} is the reverse saturation current. In conventional intensity-based imagers, the sub-threshold current is drawn by a photodiode, which has a linear current output in relation to incident light intensity [4]. A cascode transistor is implemented to increase the impedance, increasing the output voltage swing. This paper builds on the same concept to measure the time-of-flight in a compressive manner, porting the design from the intensity domain to the time domain to achieve a high degree of accuracy. The logarithmic voltage is generated by creating a stepped ramp current mirror (vis-a-vis Fig. 1(b)). From eq. 1, it can be concluded that such a current will provide a logarithmic output (as is reflected in Fig.2).
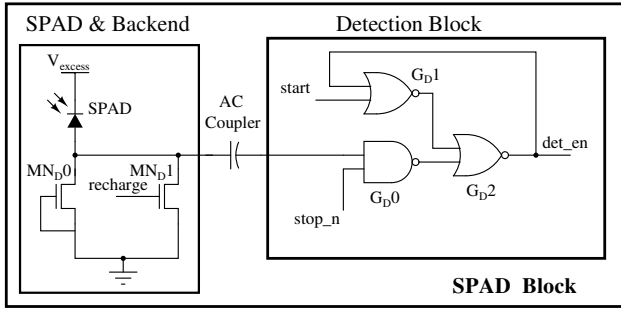
Fig. 3: SPAD block with Enable Generation [15]

II. COMPRESSIVE PIXEL DESIGN

A. SPAD Back-end and Detection Block

This pixel utilises a single photon avalanche diode (SPAD) for the detection of photons. A SPAD on the incidence of a single photon goes into the breakdown region, thus providing a large current [16]- [17]. As shown in Fig.3, a passive quench circuit is implemented with MN_D0, providing a large resistance. An active recharge circuit is created by MN_D1. A similar but modified enable generation logic, presented in [15], is implemented. The circuit generates an active-high enable 'det_en', which starts from the signal 'start'. 'det_en' goes back to low again at either the 'stop' signal or when a photon is detected, whichever event occurs first. Thus, the gap between the start and stop signals is the full-scale range of the system, and the ' det_en' signal's pulse width gives us the total ToF.
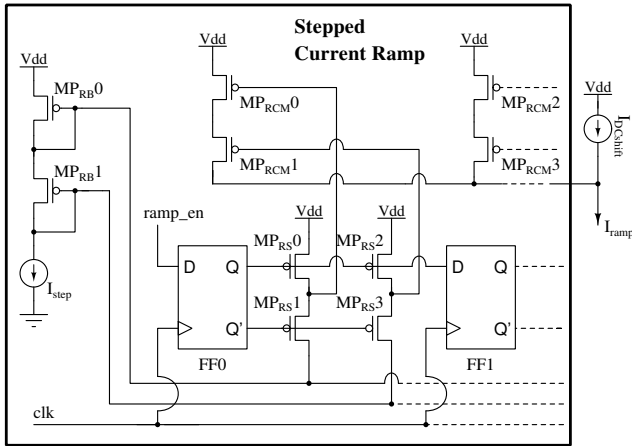
B. Time Dependent Ramp Current Mirror

Fig. 4: Stepped Ramp Current Source

A linearly time-dependent current is generated by a series of (cascode) current mirrors, which are biased sequentially using flip-flops (Fig. 4). This biased flip-flop structure creates a stepped current ramp. The biasing current used is the current increment of each step of the current ramp. $MP_{RB}0$ and $MP_{RB}1$ provide the bias voltage as cascoded current sources. Each of the flip-flops of the shift register bias their corresponding current mirroring transistors ($MP_{RCM}0$ and $MP_{RCM}1$ for FF0, $MP_{RCM}2$ and $MP_{RCM}3$ for FF1 and so on) so that they are
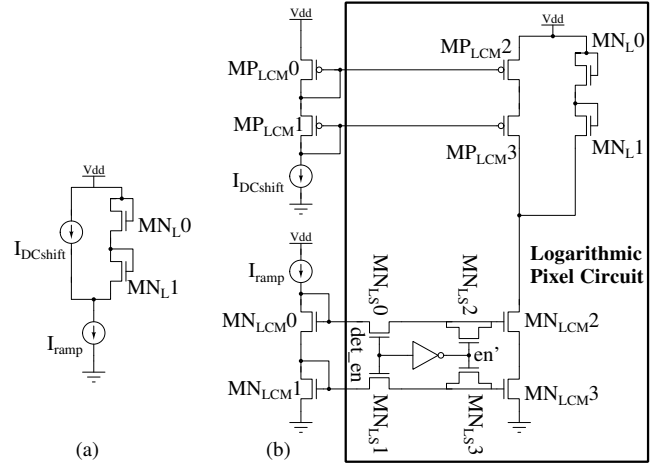
off when the flip-flop is low and are biased when the flip-flop is active. This results in a sequentially increasing number of mirroring transistors being biased as the shift register operates, creating a stepped current ramp. A larger number of biasing flip flops can be used to get a smooth ramp, reducing the significance of the stepping. As this circuit is present at either the chip level or the column level, it can be made large for better performance without any impact on the pixel-pitch (and thus the fill-factor). An always-on current of $300nA$ is added to keep the subsequent circuit in saturation. This creates a shifted current ramp that varies linearly from $300nA$ to $4\mu A$. The last flip-flop turning on also signifies the end of the logarithmic range. Thus the output signal 'log_end' of the last flip-flop (FF_n) is taken as a start signal for the subsequent linear block.

Fig. 5: (a) Model of Logarithmic Pixel (b) Log Pixel Implementation

C. Logarithmic Circuit

The logarithmic voltage is a function of the current through MN_L0. It remains in the subthreshold region for currents up to 4 μA, providing an output swing of $310mV$ (Fig. 2). As the initial current has a very large output swing, it is essential to utilise this range. This becomes a problem as current mirrors can't be in saturation for near-zero current values. Thus a differential approach is considered. A small constant current '$I_{DCshift}$' is supplied to the pixel, which the shifted ramp current dissipates (Fig. 5(a)). The excess current provided (by PMOS current mirror) is made to match the minimum of the current ramp mirror as closely as possible.

The differential design is implemented using cascode current mirrors (Fig. 5(b)). $MP_{LCM}0$ and $MP_{LCM}1$ constitute a cascode current mirror for providing the '$I_{DCshift}$' mirrored in each pixel (by $MP_{LCM}2$ and $MP_{LCM}3$). A switched cascode current mirror mirrors the ramp current (mirrored by $MN_{LCM}0$, $MN_{LCM}1$ at $MN_{LCM}2$, $MN_{LCM}3$). The switches $MN_{LS}0$ and $MN_{LS}1$ are low leakage (high V_T) NMOS switches. When the enable is high, the ramp current is mirrored; thus, the voltage generated changes logarithmically. As soon as the enable goes low, the switches turn off, disabling any change to the gate
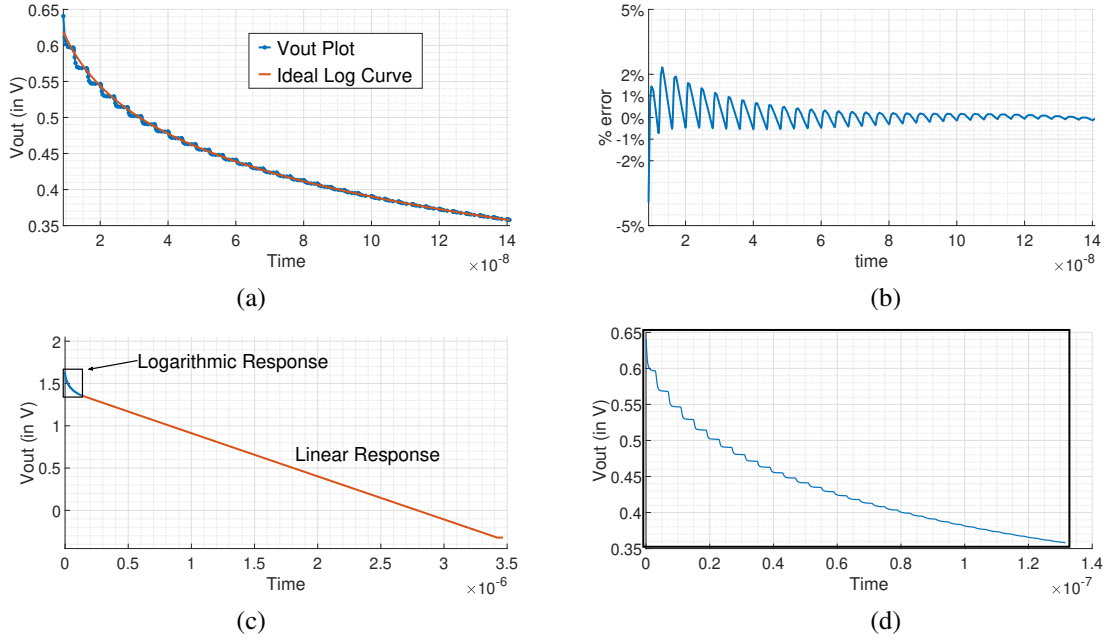
Fig. 6: (a) Achieved log response and ideal log curve (b) Percent error from ideal (c) Full-scale voltage response of pixel (d) Expanded view of logarithmic response

voltages of $MN_{LCM}2$ and $MN_{LCM}3$. The gate capacitance thus stores the bias voltage of the moment when the enable goes low. Thus the mirrored current (and consequently the logarithmic voltage) holds its value as it was at the instant of the 'det_en' signal going low. Dummy transistors $MN_{LS}2$ and $MN_{LS}3$ are implemented to mitigate the issue of the bias changing due to clock feed-through.
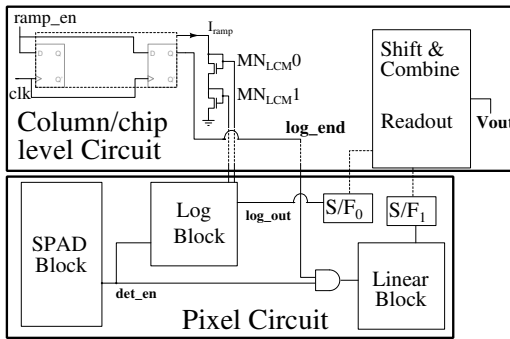
III. PIXEL OPERATION AND RESULTS

Fig. 7: System Level Design of Pixel

The system-level block diagram for the proposed pixel is shown in Fig. 7. A linear TAC circuit is implemented (as the linear block) from [15] to achieve a linear response after the logarithmic response. This circuit is configured with a low current to provide a long-range (at a low resolution). This circuit is activated after the logarithmic circuit reaches its FSR. This signal 'log_end' is received when the stepped current ramp circuit's shift register overflows. The AND of 'log_end' and 'det_en' is the enable signal for the linear circuit, thus giving us a logarithmic response initially and then switching
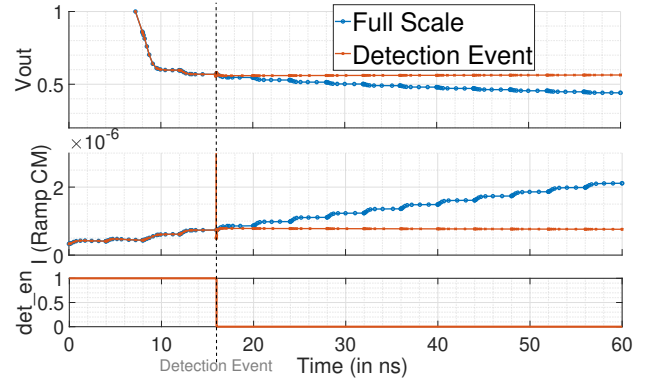

Fig. 8: Pixel response to photon incidence during logarithmic operation

to a linear response (as seen in Fig. 6(c) and Fig. 6(d)) Thus the linear circuit operates only after the FSR of the logarithmic circuit and stops when a photon is detected, or the FSR of the entire pixel is achieved, whichever occurs first. 'ramp_en' is an active high enable signal for the entire circuit, input at the shift register. Its pulse width is chosen such that the output of the shift register ('log_end') goes low after the FSR of the whole pixel. If a photon is incident when the circuit is operating in the logarithmic region, the logarithmic circuit holds the current ramp value (and consequently the voltage) at the instant of detection. The linear circuit is not activated as 'det_en' will be low hereafter (till the next 'start' signal to the SPAD block). The pixel's response to a detection event is shown in Fig. 8.

Assuming that an 8-bit ADC is implemented with the least count of 7.7 mV, the operational performance of the pixel is estimated. This pixel achieves a maximum resolution of 16

Parameter	This Work	2014 [13]	2021 [15]	2019 [7]	2011 [12]
Technique	TAC	TAC	TAC	TDC	TDC
CMOS Technology (nm)	65	130	180	180	350
Dynamic Range (FSR)	3.33 μ s	80 ns	20 ns	330 ns	160 ns
Dynamic Range (Distance)	500 m	12 m	3 m	50 m	24 m
ToF Resolution (Log/Linear)	106.6 ps-1.51/15.1 ns	NA / 93 ps	NA / 29 ps	NA / 48.8 ps	NA / 10 ps
Depth Resolution (Log/Linear)	16mm-172/226 cm	NA / 13.95 mm	NA / 11.76 mm	NA / 7.32 mm	NA / 1.5 mm
Min.Depth	16 mm	13.95 mm	11.76 mm	7.32 mm	1.5 mm

TABLE I: Comparisons with other state-of-the-art Pixels

mm (106.6 ps) in logarithmic operation, with the resolution relaxing to 172 cm (11.5 ns). The linear circuit is configured so that the pixel achieves an FSR of 500 m. The linear resolution thus achieved is 226 cm. The minimum measurable range by the pixel is 16 mm. The R^2 achieved by the logarithmic design when compared to an ideal logarithmic curve (Fig. 6(a)) is 0.9989 indicating that the achieved curve follows the ideal very closely. The maximum deviation of the logarithmic voltage from an ideal logarithmic curve is 2.5% to -0.5% as seen in Fig. 6(b).

IV. CONCLUSION

This work presents a compact compressive LiDAR pixel with multi-mode operation, which utilises a logarithmic design approach to achieve a high resolution (16 mm reducing to 172 cm) at lower ranges (upto 19.5 m) and switches to a linear circuit with a relaxed resolution (226 cm) for long-range measurement (upto 500 m).

Fig. 9: Layout of logarithmic circuit. Dimensions:8 μm×6 μm

REFERENCES

[1] B. Park et al., "A 64 × 64 SPAD-Based Indirect Time-of-Flight Image Sensor With 2-Tap Analog Pulse Counters," in IEEE Journal of Solid-State Circuits, vol. 56, no. 10, pp. 2956-2967, Oct. 2021, doi: 10.1109/JSSC.2021.3094524.

[2] C. Bamji et al., "A Review of Indirect Time-of-Flight Technologies," in IEEE Transactions on Electron Devices, vol. 69, no. 6, pp. 2779-2793, June 2022, doi: 10.1109/TED.2022.3145762.

[3] F. Villa et al., "CMOS Imager With 1024 SPADs and TDCs for Single-Photon Timing and 3-D Time-of-Flight," in IEEE Journal of Selected Topics in Quantum Electronics, vol. 20, no. 6, pp. 364-373, Nov.-Dec. 2014, Art no. 3804810, doi: 10.1109/JSTQE.2014.2342197.

[4] A. Bermak, A. Bouzerdoum and K. Eshraghian, "A high fill-factor native logarithmic pixel: Simulation, design and layout optimisation," 2000 IEEE International Symposium on Circuits and Systems (ISCAS), 2000, pp. 293-296 vol.5, doi: 10.1109/ISCAS.2000.857422.

[5] C. Anand, K. Jainwal, M. Sarkar, "A Three-Phase, One-Tap High Background Light Subtraction Time-of-Flight Camera," in IEEE Transactions on Circuits and Systems I: Regular Papers, vol. 66, no. 6, pp. 2219-2229, Jan. 2019 , doi: 10.1109/TCSI.2018.2890050.

[6] K. Jainwal, C. Anand, M. Sarkar, "1/f Noise Reduction Using In-Pixel Chopping in CMOS Image Sensors," in IEEE Solid-State Circuits Letters, vol. 1, no. 6, Jun. 2018 , doi: 10.1109/LSSC.2018.2879722.

[7] C. Zhang, S. Lindner, I. M. Antolović, J. Mata Pavia, M. Wolf and E. Charbon, "A 30-frames/s, 252×144 SPAD Flash LiDAR With 1728 Dual-Clock 48.8-ps TDCs, and Pixel-Wise Integrated Histogramming," in IEEE Journal of Solid-State Circuits, vol. 54, no. 4, pp. 1137-1151, April 2019, doi: 10.1109/JSSC.2018.2883720.

[8] S. Lindner, C. Zhang, I. M. Antolovic, M. Wolf and E. Charbon, "A 252 × 144 SPAD Pixel Flash Lidar with 1728 Dual-Clock 48.8 PS TDCs, Integrated Histogramming and 14.9-to-1 Compression in 180NM CMOS Technology," 2018 IEEE Symposium on VLSI Circuits, 2018, pp. 69-70, doi: 10.1109/VLSIC.2018.8502386.

[9] S. Kurtti, J. -P. Jansson and J. Kostamovaara, "A CMOS Receiver–TDC Chip Set for Accurate Pulsed TOF Laser Ranging," in IEEE Transactions on Instrumentation and Measurement, vol. 69, no. 5, pp. 2208-2217, May 2020, doi: 10.1109/TIM.2019.2918372.

[10] C. Anand, N. Priyadarshini, K. Jainwal, M. Sarkar, "A 125-klx Background Light Subtraction Architecture for 2-D and Time-of-Flight 3-D Cameras," in IEEE Transactions on Electron Devices, vol. 65, no. 9, pp. 3823 - 3830, Sep. 2018, doi: 10.1109/TED.2018.2860048.

[11] C. Niclass, M. Soga, H. Matsubara, M. Ogawa and M. Kagami, "A 0.18-μ m CMOS SoC for a 100-m-Range 10-Frame/s 200 × 96-Pixel Time-of-Flight Depth Sensor," in IEEE Journal of Solid-State Circuits, vol. 49, no. 1, pp. 315-330, Jan. 2014, doi: 10.1109/JSSC.2013.2284352.

[12] Markovic, S. Bellisai and F. A. Villa, "15bit Time-to-Digital Converters with 0.9% DNLrms and 160ns FSR for single-photon imagers," 2011 7th Conference on Ph.D. Research in Microelectronics and Electronics, 2011, pp. 25-28, doi: 10.1109/PRIME.2011.5966209.

[13] L. Parmesan, N. A. W. Dutton, N. J. Calder, A. J. Holmes, L. A. Grant and R. K. Henderson, "A 9.8 μm sample and hold time to amplitude converter CMOS SPAD pixel," 2014 44th European Solid State Device Research Conference (ESSDERC), 2014, pp. 290-293, doi: 10.1109/ESSDERC.2014.6948817.

[14] M. Crotti, I. Rech and M. Ghioni, "Four Channel, 40 ps Resolution, Fully Integrated Time-to-Amplitude Converter for Time-Resolved Photon Counting," in IEEE Journal of Solid-State Circuits, vol. 47, no. 3, pp. 699-708, March 2012, doi: 10.1109/JSSC.2011.2176161.

[15] Z. Wu, Y. Xu and Z. Ma, "A Time-to-Amplitude Converter With High Impedance Switch Topology for Single-Photon Time-of-Flight Measurement," in IEEE Access, vol. 9, pp. 16672-16678, 2021, doi: 10.1109/ACCESS.2021.3053758.

[16] F. Piron, D. Morrison, M. R. Yuce and J. -M. Redouté, "A Review of Single-Photon Avalanche Diode Time-of-Flight Imaging Sensor Arrays," in IEEE Sensors Journal, vol. 21, no. 11, pp. 12654-12666, 1 June1, 2021, doi: 10.1109/JSEN.2020.3039362.

[17] D. Bronzi, F. Villa, S. Tisa, A. Tosi and F. Zappa, "SPAD Figures of Merit for Photon-Counting, Photon-Timing, and Imaging Applications: A Review," in IEEE Sensors Journal, vol. 16, no. 1, pp. 3-12, Jan.1, 2016, doi: 10.1109/JSEN.2015.2483565.

Sohail Faizan - Primary Author: Sohail Faizan is Pursuing his B.Tech. (honours) from the Department of Electrical Engineering at the Indian Institute of Technology (IIT) Bhilai. His research focuses on analogue CMOS circuits and he specialises in CMOS imagers.

Kapil Jainwal - Corresponding/Primary Author: Kapil Jainwal is with the Electrical Engineering Department of the Indian Institute of Technology (IIT) Bhilai as an Assistant Professor. He received his PhD from the Electrical Engineering Department of IIT Delhi and a Master's degree from IIT Bombay.

Minal Bisen is a Ph.D. student with the Electrical Engineering Department of the Indian Institute of Technology (IIT) Bhilai.

Nitin Khanna is with the Electrical Engineering Department of the Indian Institute of Technology (IIT) Bhilai as an Associate Professor. He received his PhD from the Purdue University and B.Tech degree from IIT Delhi.

A back-illuminated full-frame low-noise HDR 8μm, 12Mpixel, 34fps image sensor for industrial, medical and scientific applications

M. Sannino[a], C. Bauza-Alcover[a], A. Font-Garcia[a], R. Gifreu-Pons[a,], M. Gomez-Navarro[a,] O. Llados-Cos[a],
A. Molla-Garcia[a], A. Scott[a], D. Uzbelger Feldman[b], E. Simons[c], A. Birman[d], R. Turchetta[a],
A. Bofill-Petit[a*]

a: IMASENIC S.L., Pl. Tetuan 40-41, 08010 Barcelona, Spain
b: Real Time Imaging Technologies, LLC, 1723 Beverly Drive, Charlotte, NC 28207, USA
c: Blur Product Development, 260 James Jackson Ave, Cary, NC 27513, USA
d: Tower Semiconductor, Shaul Amor 20, Migdal Haemek, Israel

* Email: adria.bofill@imasenic.com

Abstract — The need for a low-noise, high-dynamic range (HDR), high-resolution, back-side illuminated (BSI) sensor with a frame rate level that allows video applications for next-generation X-ray medical, scientific, and industrial equipment has been identified. Such a sensor will allow reductions in the radiation dose received by patients in medical applications and use of the sensor in low-light conditions while increasing the image resolution. It will make possible fluoroscopy and X-ray video in some healthcare and industrial applications where it is not used presently. To this end, a BSI stitched sensor with HDR, 8μm, CDS-capable pixels has been designed. The sensor has 12M pixels and can be readout at a frame rate of 34fps. The sensor is amenable to many other industrial and scientific applications that can benefit from its 1.5 e-(rms) noise level.

I. INTRODUCTION

2D and 3D radiographs are an indispensable diagnostic tool in healthcare. Consequently, annual per-capita effective doses from radiologic and nuclear medicine procedures have increased about 2,467% between 1980 and 2017. Absorbed doses from medical and dental radiography have declined by more than 60% in recent years because of reductions in x-ray kilovoltage peaks (kVp) and the introduction of front-side-illuminated CMOS digital sensor technology, among other developments. However, effective radiation doses of cumulative medical and dental imaging examinations of patients are still too high. In addition, image resolution of the available technology is not high enough to provide the level of confidence required to avoid a significant quantity of diagnostic errors.

The range of x-ray tube currents has not been taken into consideration as a way of reducing radiation dose so far. The major challenge of decreasing current level in X-ray imaging systems is the quantum noise phenomena. However, previous studies show a significant improvement in medical image quality compared to existing equipment by using BSI sensors, pixel sizes in the range of 6μm-11μm and microlenses [1][2].

This paper presents a low-noise, high-resolution, high pixel count, BSI sensor with microlenses that will allow moving medical and dental diagnostic beyond the state-of-the art. The sensor features will also enable its use in scientific and industrial applications that require similar characteristics. For X-ray applications the sensor will be coupled to a scintillator.

Making sensors able to provide higher quality image with lower radiation levels also opens the door to the use of X-ray video (fluoroscopy) in new medical and dental procedures. Similarly, the sensor presented can be used in industrial and scientific applications that require high quality video at low light levels.

II. SENSOR OVERVIEW

The sensor presented has been manufactured with Tower Semiconductor 180nm CIS process. The top-level block diagram is shown in Figure 1. The sensor features 3000 columns x 3864 rows of pixels with a pitch of 8μm. The pixel is based on a 6T pixel architecture that allows HDR operation with Correlated Double Sampling (CDS). The pixel is described in section III. The size of the focal plane array is 24.0mm x 30.9mm, whereas the die size is 25.5mm x 37.1mm. The row-drivers are implemented with a compact design and layout to reduce the die area, as required in some medical applications.

The sensor has one pixel output line per pixel column routed towards the south-side of the sensor periphery. The pixel output lines are funnelled in the periphery through an array of Programmable Gain Amplifiers (PGA), Sample and Hold stages (SH), and column-parallel ADCs. Then, the digital data is shifted into a buffer, serialized, combined with clock alignment and

data alignment patterns and, finally, fed to the sub-LVDS output data transmitters. There are 5 sub-LVDS output data channels that can operate at a maximum speed of 1.2Gbps and 1 clock output, for systems that required a clock-synchronous interface. Power and I/O pads are located only on the south-side of the die. This allows having the perimeter of the packaged sensor to be very close to the edge of the focal plane on 3 sides of the sensor.
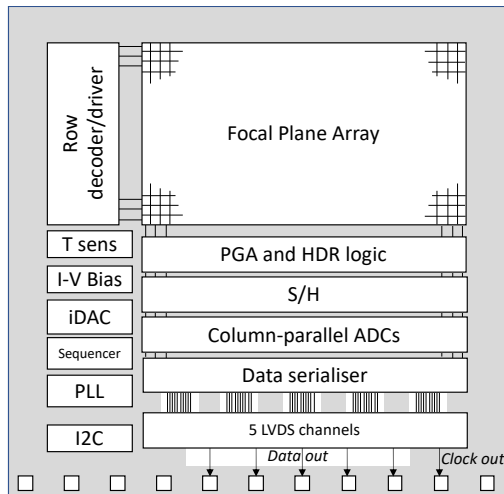

Figure 1. Sensor top-level block diagram.

The sensor includes all the internal logic to control its operation with a digital sequencer. The operation is highly configurable through an I^{2}C serial interface. In addition to the 2-wire serial bus, only power and ground lines and a low-speed (default: 25MHz) reference clock are needed from the system. The high-speed clock for the 1.2Gbps LVDS output data channels is generated on-chip with a ring-oscillator-based PLL. The output of the same clock generator goes through a clock divider block to generate the sequencer and ADC clocks. All ancillary blocks such as the programmable current biasing network with current and voltage references and current DACs and a temperature sensor are also included on-chip.

The 12M pixel sensor achieves a frame rate of 34fps in HDR mode and 40fps in dual-gain (DG) mode. These numbers are limited by the position of the IO pads (only on the south-side of the sensor) and the form factor with 30% more rows than columns. These characteristics are required by some of the target applications of the sensor. Using the same circuit blocks but with a different focal plane form factor and having pads in more than 1 side of the sensor would allow significantly higher frame rates.

The specifications are summarized in Table 1. At the time of writing, the sensor is still in manufacturing, thus the specifications given are based on design parameters and simulation results.

Table 1. Sensor specifications

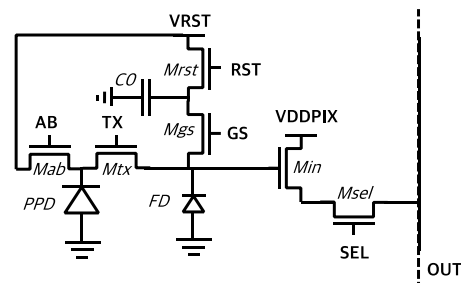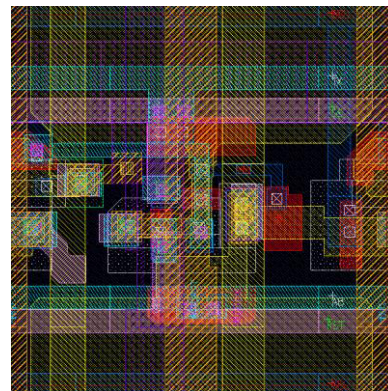| Technology	180nm CIS
Number of pixels	3000 x 3864
Effective focal plan size	24.0mm x 30.9mm
Die size	25.5mm x 37.1mm
Pixel pitch	8μm
HDR	Yes
Noise [e-rms]	1.5 e-rms
FW	≥ 65 ke-
Dynamic Range	92.7 dB
ADC on chip	12 bit
Data output interface	5 LVDS @ 1.2Gbps
Frame rate	34 fps @ full frame (HDR) 40 fps @ full frame (DG)
Stitching	Yes

Figure 2. 6T pixel schematic

Figure 3. Pixel layout.

III. PIXEL

The schematic and layout of the pixel are shown in Figure 2 and Figure 3, respectively. The pixel size is 8μm x 8μm. The pixel uses a 6T pixel architecture with lateral overflow capacitors. This enables HDR operation of the pixel. The pixel has two possible gain levels. The Mgs transistor is driven always at a low voltage level during signal integration or pixel reset reading for CDS. After the TX pulse, the signal is first read with the pixel in High Gain (HG). For this, transistor Mgs is driven with a low gate voltage. In this configuration, when the PGA further down the readout chain in the periphery is set to its maximum gain of 8

(linear), the total readout noise is 1.5e-rms. In HDR mode, after the HG reading the gate of transistor Mgs is driven to a high-voltage and the pixel is read in low gain (LG). In this configuration, the pixel has higher noise and a higher full-well (65ke).

The sensor can be configured to work in HDR mode or in dual-gain mode. In this second mode of operation the pixels are read with a fixed gain setting, either LG or HG. In this mode higher frame rate can be achieved if the pixel timings are modified such that only LG or HG reading is done after reset reading for CDS.

IV. READOUT ARCHITECTURE

The signal readout architecture in the south side of the sensor periphery is shown in Figure 4. The first stage in the readout chain following the pixel matrix is an 8-level programmable gain amplifier (PGA). In HDR mode the gain of the amplifier is adjusted automatically depending on the pixel signal strength between two pre-selected gain settings. For LG the default PGA gain is 1, whereas for HG the default gain is 8. These defaults can be modified by the sensor application through the I²C interface.

The HDR-logic circuit will set the gain bit (GBIT) signal HIGH if the output of the PGA is saturated when reading the pixel in HG. The HDR logic output is controlled by the sensor timing sequencer in dual-gain mode.

In order to achieve higher data rate, the Sample & Hold (SH) stage has two storing capacitors operated in ping-pong (SH1 and SH2). The ADC conversion of one row signal (N) is pipelined with the signal analog settling from the next row (N+1).

In HDR mode, for each row, one of the SH capacitors first samples the HG gain reading of the pixel at the PGA output. If the output of the PGA is higher than a programmable VREF_GBIT saturation limit, then, when reading the pixel in LG, the same SH capacitor will sample the LG reading (the HG sample will be lost). If the PGA is not saturated during HG read, then the LG sample is not stored in any SH capacitor and this is the reading which is ignored. The ADC only does one conversion per pixel. When the pixel reading in HG is above the saturation limit the ADC will convert the LG pixel signal. Otherwise, it will convert the HG signal. An example of this sequence is show in Figure 5. It also shows the SH stage ping-pong operation. In this example, the pixels in row N for a given column receive a strong signal that requires LG, whereas pixels in row N+1 receive a weaker signal that requires HG.

The dual-gain mode operation is shown in Figure 6. In this case, the readout chain in the periphery only receives the LG or the HG reading of the pixel as configured through the serial interface of the sensor.

The last stage in the readout chain in Figure 4 is a 12-bit incremental sigma-delta ADC. The GBIT is added to the ADC output code for each conversion to indicate the pixel and PGA gain used for this reading. The final 13bit code will be used by the system host for HDR pixel value reconstruction. The design of the ADC is similar to the one presented in [3].

V. DATA SERIALIZATON AND STREAM-OUT

The 5 readout channels in the sensor are each assigned to 600 ADCs and composed of a 600:16 multiplexer, a 16 words' buffer, a serializer and a sub-LVDS output, as shown in Figure 7.

After the ADCs have completed the conversion, the 12-bit result is stored in buffers internal to each ADC. In HDR mode, the GBIT is also sampled into the buffer. This sampling allows the pipelining of the ADC conversion with the data stream-out, as shown in Figure 5 and 6. The ADCs are subsequently accessed in groups of 16 at a time via a 600:16 multiplexer (since 600 is not a multiple of 16, the last selection includes 8 ADCs instead of 16), and their data is shifted out and stored in a 16 words' buffer. The total size of the buffer is configurable 12x16 or 13x16 depending on whether the selected mode is dual-gain or HDR, respectively. After storing the data into the buffer, it is shifted out to the serializer and ultimately streamed out by the sub-LVDS driver at 1.2 Gbps at double-data rate. The storage of data from the N+1th group of ADCs and the stream-out of the data from the Nth group are pipelined, allowing a seamless data stream, as shown in Figure 8.

VI. CONCLUSION

The use of smaller pixel sizes than is common in medical and dental sensors, a low-noise HDR pixel, high-data rate throughput, BSI technology and microlenses in the sensor presented will increase patient safety by reducing overexposure to radiation while increasing diagnostic confidence. The sensor does not only target the medical and dental markets but it can also be used in industrial and scientific applications that need low-noise and high-dynamic range operation. The sensor is stitched so other sensor sizes can be manufactured with the same mask set.

VII. REFERENCES

[1] Mistry AR, Uzbelger Feldman D, Yang J, Ryterski E. *Low dose x-ray sources and high quantum efficiency sensors: the next challenge in dental digital imaging?* Radiology Research & Practice. Vol. 2014, Article ID 543524, 7 pages, 2014. doi:10.1155/2014/543524

[2] Uzbelger Feldman D, Yang J. *Milliamperes settings and image noise reduction through back illumination and adaptive filtration.* 66th Annual Session of American Academy of Oral & Maxillofacial Radiology, Indianapolis, IN, 2015.

[3] M. Sannino, A. Bofill-Petit, G. Pinaroli and R. Turchetta, *A high dynamic range, 1.9 Mpixel CMOS image sensor for X-ray imaging with in-pixel charge binning and column parallel ADC*, IISW 2019.
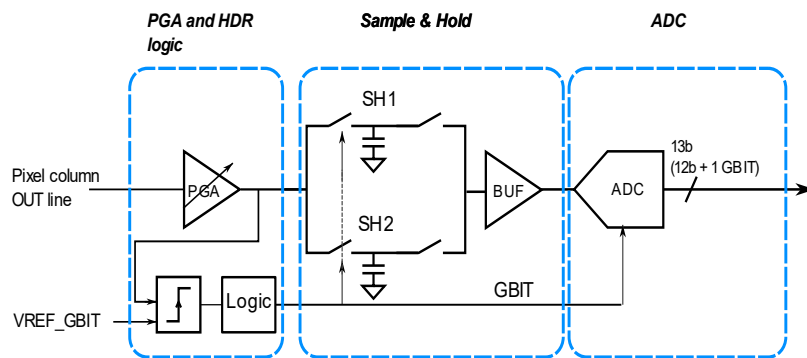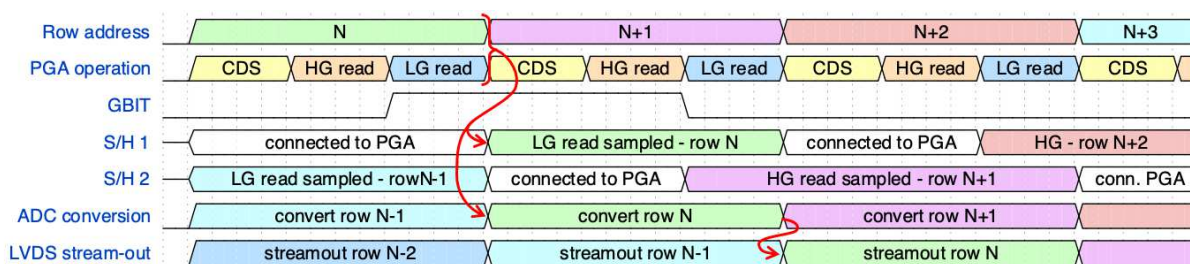
Figure 4. Readout architecture
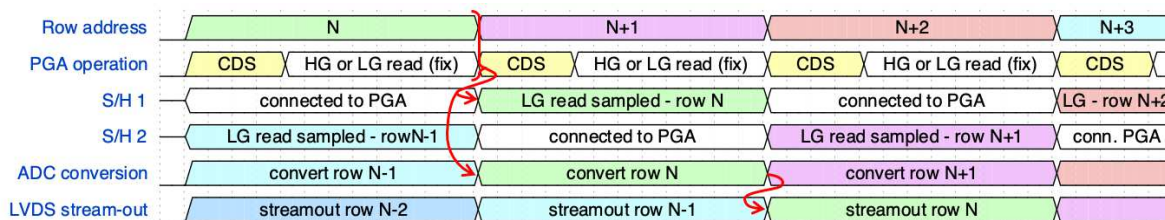

Figure 5. HDR readout operation.

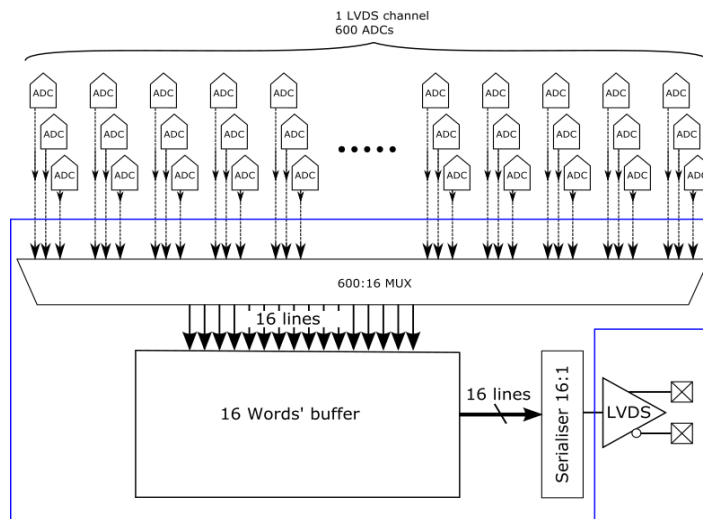*Figure 6. Dual-gain readout operation.*

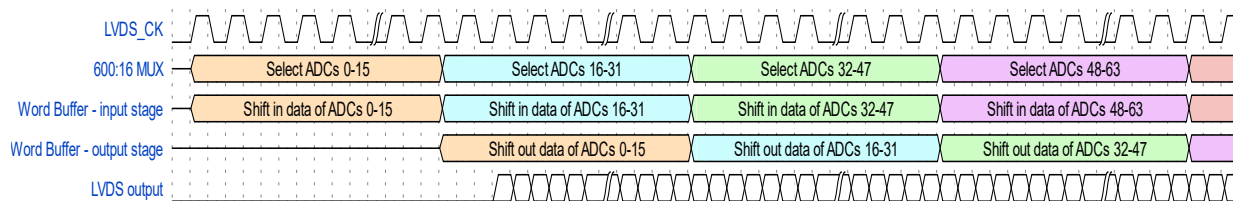*Figure 7. Data serialization and streamout architecture*

Figure 8. Data serialization operation

Front- / Backside Illuminated Low Noise Embedded CCD image sensor with Multi Level Anti Blooming functionality

Olaf Schrey
Optical Systems
Fraunhofer Institute for Micro-
electronic Circuits and Systems IMS
Duisburg, Germany
olaf.schrey@ims.fraunhofer.de

Denis Piechaczek
Optical Systems
Fraunhofer Institute for Micro-
electronic Circuits and Systems IMS
Duisburg, Germany
denis.piechaczek@ims.fraunhofer.de

Manuel Ligges
Optical Systems
Fraunhofer Institute for Micro-
electronic Circuits and Systems IMS
Duisburg, Germany
manuel.ligges@ims.fraunhofer.de

Bedrich J. Hosticka
Optical Systems
Fraunhofer Institute for Micro-
electronic Circuits and Systems IMS
Duisburg, Germany
bedrich.hosticka@ims.fraunhofer.de

Abstract— **This paper presents a 320 columns x 128 lines Time Delay and Integration (TDI) image sensor with an embedded charge-coupled device (eCCD) structure fabricated in a 0.35 μm high voltage CMOS process. The paper focuses on the noise characteristics of the sensor analog readout chain and presents an analytical model for effective noise reduction down to an equivalent input noise (ENC) of 20 e⁻. The TDI sensor utilizes bidirectional charge shifting and additional excess charge draining by the pixel reset transistor. These techniques provide an effective anti blooming capability, which together with the optimized noise characteristics offers a dynamic range of 13 bit.**

Keywords—noise; conversion gain; correlated double sampling; time-delayed integration; charge-coupled device; responsivity; photon transfer curve; signal-to-noise ratio; blooming; backside illumination; frontside illumination;

I. INTRODUCTION

Commencing with a brief illustration of the target application specification, the paper introduces the eCCD cell and the analog readout path constituting the key building blocks of the TDI sensor. The subsequent sections address the bidirectional charge shift implementation followed by a noise model of the full analog signal path emphasizing the effectiveness of correlated double sampling (CDS). The paper concludes with a presentation of measurement results covering a Photon Transfer Curve (PTC) method based noise characterization and the bidirectional charge shift functionality showing its impact on TDI-linearity.

II. EARTH OBSERVATION BOUNDARY CONDITIONS

Satellite-based earth observation systems are required to detect and distinguish low reflectivity objects under high background irradiance conditions (typically sunlight). Imaging systems utilizing CCDs thus have to offer high intrascene dynamic range (DR), high signal to noise ratio (SNR) and optimum scene contrast resolution capability [1]. High DR is given by maximizing the full well capacity (FWC) of the CCD cell. High SNR is provided by operating the CCD sensor in time-delay integration (TDI) mode. Sensor contrast resolution is directly related to its full well capacity (FWC) and disturbing charge (dark current, blooming charge). In order to cover a large viewing area, earth observation satellites are operating in the exosphere, at orbital altitudes of approximately 600 km. Resolving ground objects of less than a buildings size requires a ground speed distance (GSD) of less than 1 m. Assuming a focal length of 8 m, a pixel pitch of 7 μm would result (Fig. 1). Gravitational and centripetal force equilibrium requires a satellite ground speed (GS) of 7.1 km/s. For the TDI sensor to be in synchronization with its ground speed, the TDI line frequency f_{line} calculates according to (1).

$$f_{line} = \frac{GS}{GSD} = \frac{7,100 \, ^{m}/_{S}}{0.5 \, m} = 14.1 \, kHz \qquad (1)$$
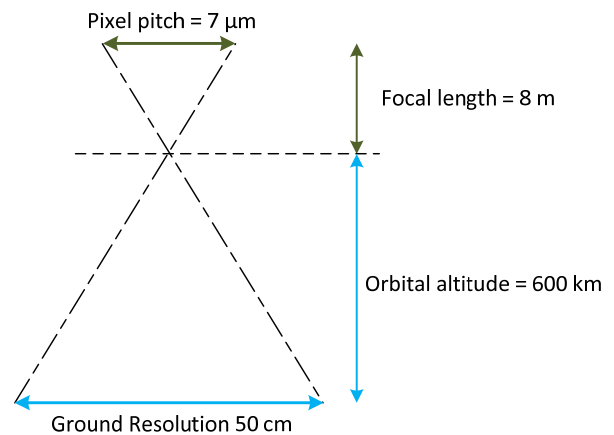

Fig. 1. Ground resolution (600 km orbit)

The targeted TDI sensor performance parameters are summarized in TABLE I. and further discussed in the following sections.

TABLE I.   TDI SENSOR KEY PERFORMANCE PARAMETERS

Parameter	*Variable*	*Value*	*Unit*
Dynamic Range	DR	13	bit
Pixel Pitch	p	7	μm
Line rate	f_{line}	14.1	kHz

III. TDI SENSOR BUILDINGN BLOCKS

The pixel geometry is given by the input requirements GSD and orbital altitude. Starting with the eCCD cell design, the following subsections introduce the circuit solution concepts.

A. CCD cell design

The high voltage regime of our 0.35 µm eCCD process allows a maximum potential difference of 16 V. For a square 7 µm x 7 µm pixel, a reduced eCCD column is numerically modeled using Synopys TCAD™. Fig. 2 depicts the potential profile when the pixel carries a full well capacity (FWC) of 168 ke⁻, ensuring that the photocharge does not get into contact with the surface.
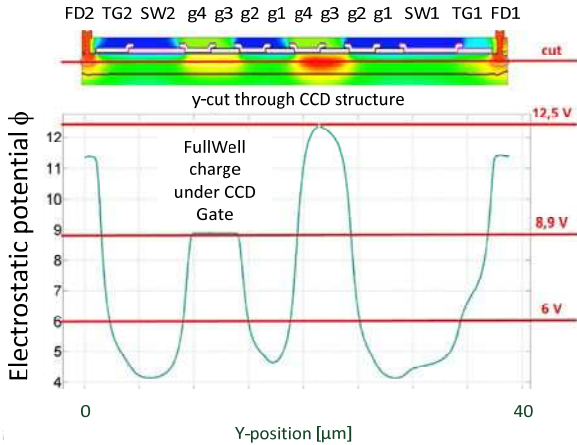
Fig. 2. CCD column potential profile at FWC. g1-4 denotes the individual 4-phase gates within a single pixel. TG1/2, SW1/2, FD1/2 denote the transfer gates, summing wells and floating diffusions adressable for bidirectional charge shift, respectively

Utilizing a 4-phase trapezoidal shaped TDI charge shift clock, the charge distributes under 2 gates at any time within a line cycle (Fig. 3).
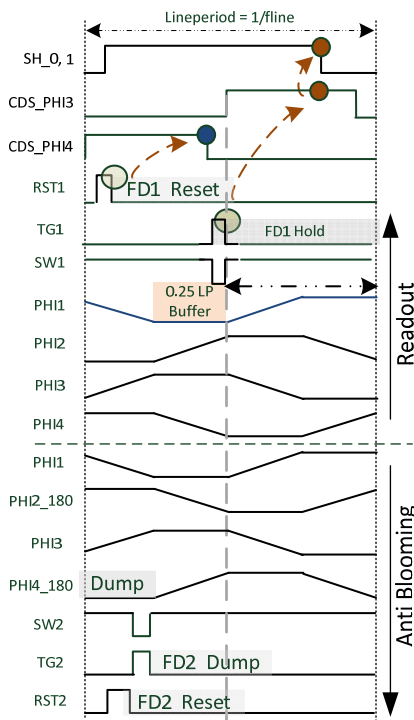

Fig. 3. Bidirectional charge shift in conjunction with CDS

CCD gates g1…g4 are fed with shift sequence PHI1 … PHI4. Excess charge (i. e. not belonging to the programmed TDI depth) is drained to the opposite side (either FD1, or FD2) by applying g2, g4 with the respective inverted signal PHI1-180, PHI4-180. Bidirectional TDI shift is employed by swapping readout / dump side through selection of summing wells (SW1/2), transfer gates (TG1/2) and floating diffusions (FD1/2). The voltage levels are set according to TABLE II.

TABLE II. ECCD GATE OPERATING LEVELS

Node / Gate	Vmin [V]	Vmax [V]	Remark
Floating Diffusion (FD1, FD2)	0 … 8	12	Reset voltage, Fill & Spill
TDI Gate (g1 … g4)	-3.5	+6.5	Charge shift
Summing Well (SW1, SW2)	-3.5	+6.5	Charge collect and buffer
Transfer Gate (TG1, TG2)	-4.5	-3.5	Charge transfer to FD

Fig. 4 depicts the column readout circuit of the eCCD element. Our eCCD process offers full CMOS integration together with the 16 V high voltage domain of the CCD structure.
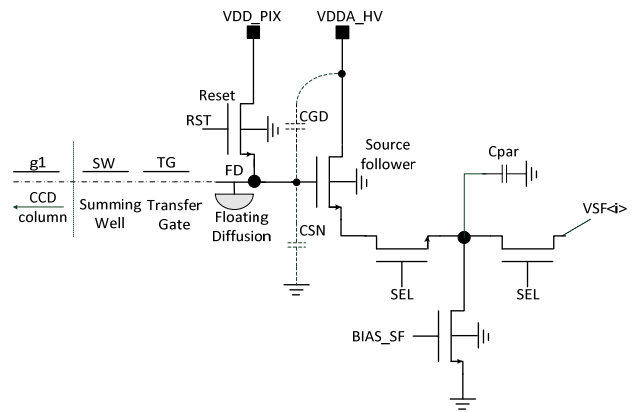

Fig. 4. eCCD column readout circuit (at top and bottom of each column)

Upon transfer from SW to FD via TG, the accumulated photocharge discharges the previously reset FD sense node capacitance (CSN). The resulting voltage (V_FD) is buffered by a high voltage source follower and fed to a correlated double sampling readout stage (CDS, discussed in the next section).
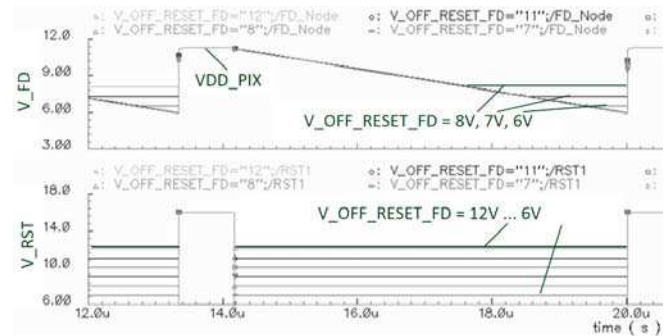

Fig. 5. Anti blooming function of pixel reset transistor. Top figure shows FD node potential, Bottom figure shows reset transistor gate voltage

Depending on the applied "V_OFF_RESET_FD"-Level, the reset transistor acts as an anti blooming gate, whenever the

FD potential becomes less than V_OFF_RESET_FD minus the reset threshold voltage. Fig. 5 shows a parametric simulation with the FD node (top waveform) being "pinned" depending on the OFF-Level of the reset transistor.

B. CDS Readout design

CDS operation is given by sampling a "reset" value onto a sampling capacitance (CS), followed by a "signal" value with the differential charge finally being transferred onto a feedback capacitance (CF) according to (2).

$$VCDS = \frac{C_S}{C_F} \cdot \left(VSF_{reset} - VSF_{signal} \right) + Vrefcds \qquad (2)$$

In the context of the eCCD, CDS is employed by resetting FD1 (RST1 being "HIGH" & CDS_PHI4 active) prior to charge transfer (TG1 being "HIGH" & CDS_PHI3 active, Fig. 3, Fig. 6).
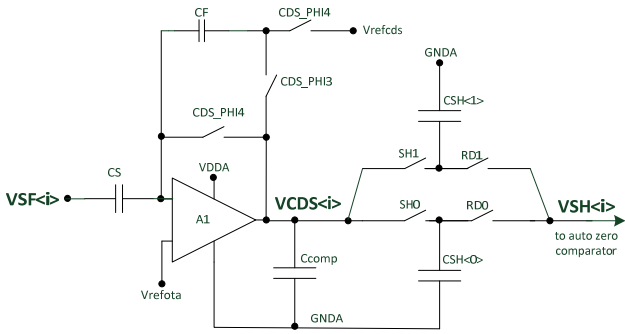
Fig. 6. Correlated double sampling (CDS) stage with Sample & Hold

CDS effectively cancels out the thermal noise of the pixel reset transistor, which is dominant for a carefully designed analog readout path [2]. According to TABLE I. , the intrascene DR is given by the ratio of FWC and equivalent noise charge (ENC) at the input node (CSN). The ENC is gained by referring the total noise voltage $V_{n,out,tot}$ at the imaging system´s output (which can be measured by PTC method [3]) to the total gain of the analog readout path, according to (3).

$$ENC = \frac{V_{n,out,tot}}{C_S/C_F \cdot ASF \cdot CG} \qquad (3)$$

Conversion gain ($CG = \frac{q_{e/e^-}}{CSN}$) and source follower attenuation (ASF) are constrained by FD layout, parasitic wiring and source follower gate area. The sense node capacitance constitutes of about 1/3 FD well capacitance, 1/3 parasitic wiring and 1/3 SF gate area, leading to a total value of 20 fF, i.e. a conversion gain (CG) of 8 $\frac{\mu V}{e^-}$.

$$V_{n,dark} = \sqrt{\frac{TDI \cdot t_{int} \cdot q_{e^-} \cdot I_{dark} \cdot ASF^2 \cdot \left(C_S/C_F \right)^2}{CSN^2}} \qquad (4)$$

Given a FWC of 168 ke⁻ and a DR of 13 bits or 8192 gray value steps, the input referred noise has to be less than 21 e⁻, hence CDS DC-Gain CS/CF has to be derived properly. We chose to employ an analytical noise model, giving the opportunity to evaluate the readout path´s noise contributors

individually. Starting with the CCD detecting element, the dark current shot noise is modeled according to (4). High TDI depth´s and slow line rates ("TDI", "t_{int}" in (4)) are driving the dark noise as does the temperature dependent dark current I_{dark}. Dark current characteristic is a matter of careful process definition and control. The temporal noise of the reset transistor is corrected by CDS and given by (5).

$$V_{n,\frac{kt}{C},reset} = \sqrt{\frac{KT \cdot ASF^2 \cdot \left(C_S/C_F \right)^2}{CSN + \frac{2}{3}CGD}} \qquad (5)$$

Besides from the sense node capacitance CSN itself, the gate drain capacitance CGD of the source follower is band limiting in favor of kT/C noise reduction (cf. Fig. 4), which also holds for the reset transistor´s partition noise (6).

$$V_{n,part,reset} = \sqrt{\frac{6 \cdot KT \cdot CRES \cdot ASF^2 \cdot \left(C_S/C_F \right)^2}{\pi^2 \cdot \left(CSN + \frac{2}{3}CGD \right)^2}} \qquad (6)$$

Partitioning noise scales with the reset gate area (CRES), while also being cancelled by CDS. The active readout chain elements - source follower (7) and CDS stage (8) - impact the thermal noise with their bias settings ratio $\frac{g_{m,Bias}}{g_{m,SF}}$, excess noise factor g and parasitic column capacitance C_{par}.

$$V_{n,therm,SF} = \sqrt{\left(1 + \frac{7}{6} \cdot \frac{g_{m,Bias}}{g_{m,SF}} \right) \cdot \frac{g \cdot KT \cdot ASF^2 \cdot \left(C_S/C_F \right)^2}{CL}} \qquad (7)$$

Totaling the thermal and kT/C noise of the CDS amplifier and the readout capacitances yields (8).

$$V_{n,CDS} = \sqrt{\frac{2 \cdot g \cdot KT \cdot \left(1 + \frac{C_S + C_{par}}{C_F} \right)^2}{C_S + C_{par} + CH \cdot \left(1 + \frac{C_S + C_{par}}{C_F} \right)} + \frac{KT \cdot (C_S + C_F)}{C_F^2}} \qquad (8)$$

The total rms noise voltage $V_{n,out,tot}$ is dominated by $V_{n,therm,SF}$ and $V_{n,CDS}$ with $V_{n,dark}$ gaining relevance for long exposure times. According to (3), the input referred noise scales with the CDS gain ratio CS/CF, with the CDS compensation capacitance $CH = C_{comp} + CSH$ (cf. Fig. 6).

TABLE III. DESIGN PARAMETERS FOR NOISE ANALYSIS

Parameter (Condition)	Variable	Value	Unit
Dark current	I_{dark}	2	fA
Sense node capacitance	CSN	20	fF
SF attenuation	ASF	0.68	
SF gate drain capacitance	CGD	25	fF
Reset gate capacitance	CRES	6.6	fF
Bias / SF ratio	$\frac{g_{m,Bias}}{g_{m,SF}}$	0.56	
Excess noise factor	g	5/3	
Parasitic column capacitance	C_{par}	2.14	pF
TDI stages	TDI	1, 32, 64, 128	

TABLE III summarizes the associated values. For a set CDS feedback capacitance of CF = 1pF, the input sampling capacitance CS is varied by +/- 10 % from its nominal value (700 fF). The influence on the ENC is shown in Fig. 7.
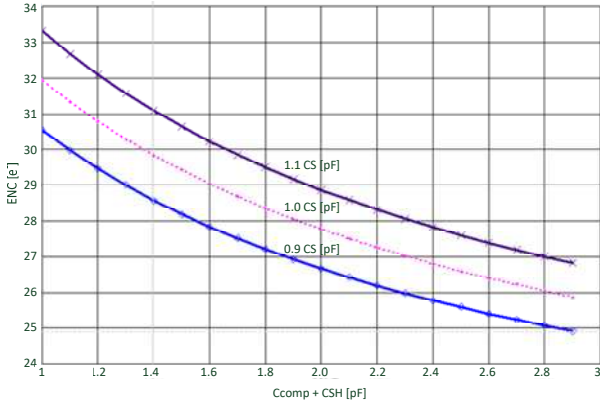

Fig. 7.   Input noise vs. total CDS compensation capacitance CH

Above 3pF the ENC is effectively reduced to less than 20 e⁻.

IV. 320 X 128 TDI ECCD SENSOR
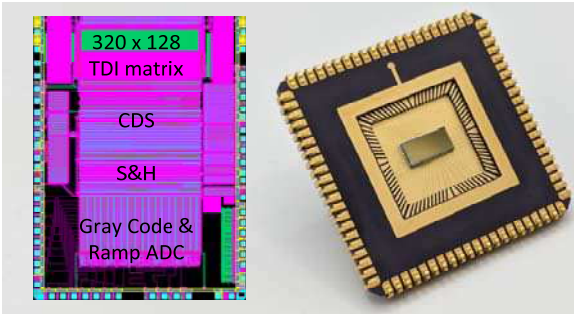
The realized eCCD image sensor is depicted in Fig. 8.

Fig. 8.   320 x 128 TDI eCCD sensor (left: layout, right: chip photo)

A. PTC characterization

The PTC characterization yields a mean camera gain of $K = 6.5 \cdot 10^{-3} \frac{DN}{e^-}$, as shown in Fig. 9.
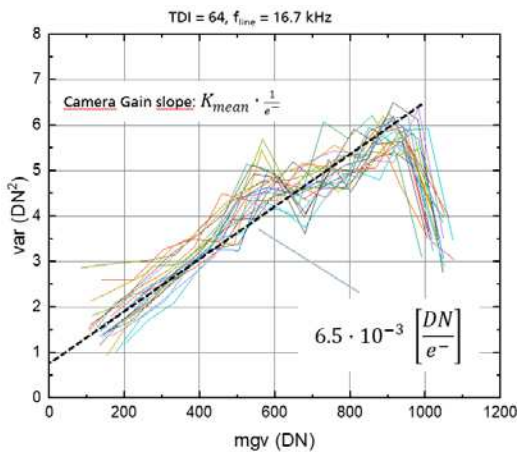
Fig. 9.   PTC characterization

The analog readout chain camera gain K is represented by (9).

$$K\left[\frac{DN}{e^-}\right] = q_{ADC}\left[\frac{DN}{V}\right] \cdot \frac{c_S}{C_F} \cdot ASF \cdot CG\left[\frac{V}{e^-}\right] \qquad (9)$$

Rearranging and substituting CG for CSN (3) yields (10).

$$C_{SN}\left[\frac{As}{V}\right] = q_{ADC}\left[\frac{DN}{V}\right] \cdot \frac{c_S}{C_F} \cdot ASF \cdot \frac{qe^-}{K\left[\frac{DN}{e^-}\right]} \qquad (10)$$

Ramp ADC conversion gain is set to $q_{ADC} = 1695 \frac{DN}{V}$. Together with the design parameters in TABLE III. , this confirms the targeted sense node capacitance of 20 fF. The dark noise is less than 1 DN, confirming the noise model.

B. Anti Blooming

Anti blooming measures (AB) incorporate AB functionality of the pixel reset in conjunction with bidirectional charge shift. Sole PTC measurement is insufficient here, since it does not distinguish between blooming and object charges. Only the determination of pixel-to-pixel modulation reveals the evidence of image quality degradation if no anti blooming measures are taken [4]. The PTC measurement characteristics depicted in Fig. 10 prove the correct AB functionality, since the TDI-depth normalized curves yield an equal slope, i.e. a – by definition - TDI-independent responsivity of $R = 3.68 \cdot 10^{-3} \frac{DN}{photon}$.
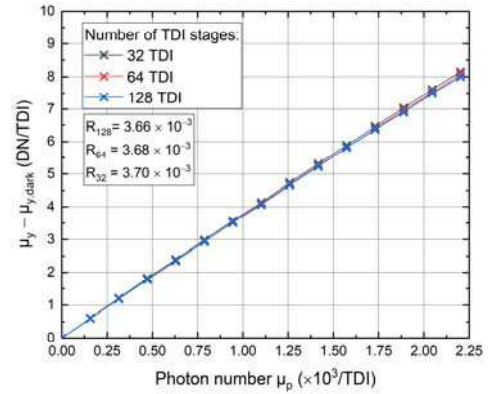

Fig. 10. TDI-depth normalized responsivity

CONCLUSION

An embedded CCD image sensor with high dynamic range, low noise and enhanced anti blooming functionality has been presented. Current development projects are based on the scalability of the low noise analog readout column element for large eCCD detectors (cf. Fig. 8). Bidirectional anti blooming clocking enables backside illuminated imaging in multi spectral earth observation applications, due to its high immunity against disturbing photocharges generated by high background irradiance.

REFERENCES

[1] Eckardt, A.; Glaesener, S.; Reulke, R.; Sengebusch, K.; Zender, B. Status of the next generation CMOS-TDI detector for high-resolution imaging. In Proceedings of the Earth Observing Systems XXIV. International Society for Optics and Photonics, 2019, Vol. 11127.

[2] Y. Degerli, F. Lavernhe, P. Magnan and J. A. Farre, "Analysis and reduction of signal readout circuitry temporal noise in CMOS image sensors for low-light levels," in IEEE Transactions on Electron Devices, vol. 47, no. 5, pp. 949-962, May 2000, doi: 10.1109/16.841226.

[3] Association, E.M.V.; et al. EMVA standard 1288, standard for characterization of image sensors and cameras. Release 2010, 3, 29.

[4] Piechaczek, D.S.; Schrey, O.; Ligges, M.; Hosticka, B.; Kokozinski, R. Anti-Blooming Clocking for Time-Delay Integration CCDs. Sensors 2022, 22, 7520. https:// doi.org/10.3390/s22197520

High Dynamic Range Pinned Photodiode Pixel with Floating Gate Readout and Dual Gain

Konstantin D. Stefanov and Martin J. Prest

Centre for Electronic Imaging (CEI), The Open University, Walton Hall, Milton Keynes MK7 6AA, United Kingdom

E-mail: Konstantin.Stefanov@open.ac.uk, Tel.: +44 1908 332116

Abstract—**This paper presents a pixel based on the pinned photodiode (PPD) with high dynamic range achieved via in-pixel dual conversion gain. The pixel operates with a single exposure and a single charge transfer out of the PPD. The signal charge is first converted to voltage non-destructively with low gain using capacitive coupling to a floating gate. A second conversion with high gain follows at a *pn* junction-based sense node after another charge transfer. An increased dynamic range is achieved due to the sensing of the same charge with two different conversion gains. The results from a prototype 10 μm pitch pixel, manufactured in a 180 nm CMOS image sensor process, demonstrate conversion gain ratio of 3:1, dynamic range of 93.5 dB, 2.4 e⁻ RMS readout noise, and negligible image lag. The pixel can operate in global shutter mode with the same low noise as in rolling shutter due to the intermediate signal storage under the floating gate.**

Keywords—**CMOS image sensor, pinned photodiode, high dynamic range, dual conversion gain**

I. INTRODUCTION

CMOS image sensors with high dynamic range (HDR) are finding use in many applications, such as automotive, surveillance, industrial, and scientific. Among the huge variety of existing HDR methods, those using a single exposure are preferred when motion artifacts must be minimized, for example in automotive imaging [1].

One of the most widely used methods to boost the dynamic range (DR) is to use column-level amplifiers with dual gain [2]. Other types of single exposure HDR imagers implement multiple signal readout paths within the pixel, such as dual photodiodes [3] or multiple conversion gains [4][5]. In-pixel signal storage with lateral overflow integration capacitor (LOFIC) [6][7] offers some of the highest DR because the maximum signal is not limited by the capacity of the photodiode.

Virtually all HDR CMOS image sensors use the pinned photodiode (PPD) as the photosensitive element due to its low dark current and readout noise. The maximum output signal, commonly known as the full well capacity (FWC), is often not limited by the charge capacity of the PPD, but by the available voltage span at the sense node. This is particularly true for larger pixels above approximately 5 μm pitch. For a typical area charge capacity of 4 ke⁻/μm², a PPD on 5 μm pitch and 60% fill factor can hold up to 60 ke⁻. On the other hand, the voltage span at the sense node cannot be much larger than 1.5 V, which for a modest conversion gain (charge to voltage factor, CVF) of 50 μV/e⁻ corresponds to 30 ke⁻. Increasing the CVF further limits the amount of charge that can be converted, and although the DR may be improved due to the lower readout noise, a DR significantly higher than 80 dB is difficult to achieve in this way.

The presented development aims to keep the readout noise low while allowing a much larger part of the charge stored in the PPD to be converted to voltage before signal saturation is reached, thus increasing the dynamic range.

II. PRINCIPLES OF OPERATION

In the proposed pixel the signal charge is converted to voltage twice following a single exposure and a single transfer out of the PPD: first on a floating gate with low CVF (low gain), and then a second time on a sense node with high CVF (high gain). The two voltage output signals are read out consecutively.

Figure 1 shows a simplified schematic diagram of the pixel [8]. A normal PPD is used as the photosensitive element. Following the transfer gate (TG), the charge passes under a floating sense gate (SG) and an output gate (OG) before reaching the sense node (SN). The MOSFET M1 is used to reset SG to the reference voltage VREF. M2 connects SN to SG so that both can be reset to VREF, and also to allow readout using only one source follower, the M3. M4 is the row select switch. Global reset functionality is provided via the GRST gate, enabling the pixel to operate in global shutter (GS) mode.

An optional *n*-type implant is included under the gates SG and OG with the purpose to move the potential peak away from the $Si-SiO_2$ interface like in buried channel charge coupled devices (CCD). This is intended to reduce the interaction of the charge with the interface states and to improve the image lag.

The diagram in Fig. 2(a) depicts the potentials in the pixel at the end of the signal integration period, and after both the SG and the SN have been reset to VREF by turning M1 and M2 on. With M1 turned off and M2 kept on, the charge is transferred out of the PPD with the help of a voltage pulse applied to the transfer gate (Fig. 2(b)). The charge is stored under the sense gate while the OG is biased at 0 V to create a potential barrier to the sense node. The signal-induced voltage step on SG is read out via M2, M3 and M4 in Fig. 2(c). This readout path has low CVF due to the large area of the SG and its high capacitance to substrate, and also due to the junction capacitances of M1, M2 and the SN which are all connected in parallel with the SG.
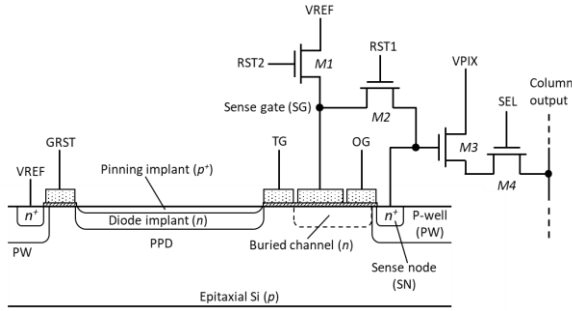
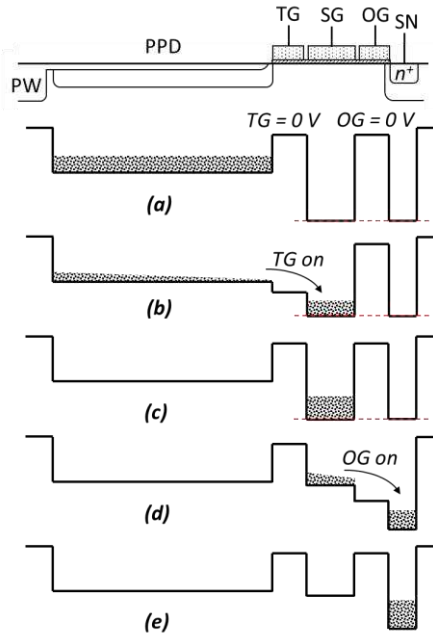Fig. 1. Schematic diagram of the dual gain pixel.

Fig. 2. Potential diagrams in: (a) signal integration with SN and SG under reset; (b) charge transfer from the PPD; (c) non-destructive signal readout at the SG; (d) charge transfer to the SN; (e) signal readout at the SN.

The first readout is non-destructive because the photogenerated signal is kept as a charge packet and is not affected by the process. Following that, the charge is transferred to the sense node for the normal charge-to-voltage conversion on a *pn* junction. First, the SN is reset to the voltage VREF by simultaneously turning on M1 an M2, while the charge is kept under SG. Following that, the voltage VREF is lowered while M1 is on, M2 is off, and OG is biased in a way to create a potential gradient towards the SN, as shown in Fig. 2(d). The charge reaches the SN and the signal is read out in Fig. 2(e). This second readout path has high conversion gain because the capacitance of the sense node is small and M2 is turned off.

III. DESIGN

A prototype pixel with the layout shown in Fig. 3 was included in a test chip together with normal PPD (4T) reference pixels. All were laid on a 10 μm pitch, with each type occupying a 128×512-pixel sub-array.

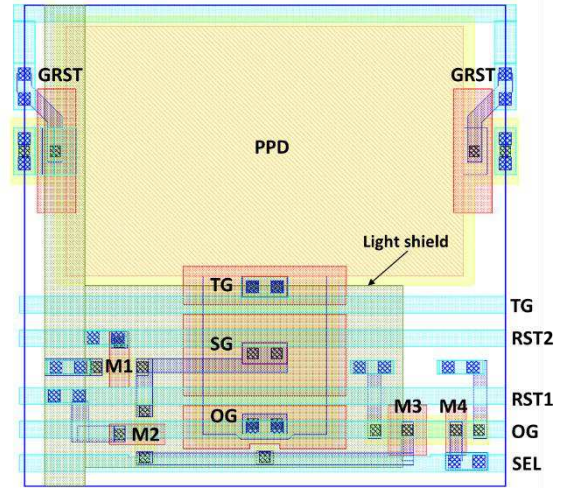Fig. 3. Layout of the dual gain pixel.

The device was manufactured in a 180 nm CMOS image sensor process using 5 μm thick, *p*-type epitaxial wafers with resistivity of 8 Ω.cmThe estimated charge storage capacity of the PPD is 170 ke⁻. From TCAD simulations of the charge coupling to the SG, and after including all the capacitances from the layout, the expected CVF for the low and high gain paths was calculated as 13.7 μV/e⁻ and 54.2 μV/e⁻, respectively.

The metal track providing the source follower supply VPIX has been enlarged to form a light shield over the SG and the OG.

A distinct advantage of the proposed pixel design is that no changes to the manufacturing process are required if the additional buried channel under SG and OG is not implemented.

IV. CHARACTERISATION

The performance of the pixel was characterized in rolling shutter mode readout using the control signals shown in Fig. 4. In addition to the signals required to achieve the operation described in Section II, an optional charge clear was added before the PPD charge transfer. This is used to remove the charge collected under the SG during integration and is identical to the transfer SG-OG-SN shown in Fig. 2(c)-(e).

The sample & hold reset (SHR) is used to store the reset samples in the column circuitry, and SHS does this for the signal samples. Correlated double sampling (CDS) is implemented by storing the output signal voltages in capacitor pairs before and after the corresponding charge transfer, thus eliminating the reset noise for both conversion gains.

Figure 5 shows the photoresponse under visible illumination for the following operating conditions: VREF1 = 2.9 V, VREF2 = 1.0 V, VOG = 1.5 V, VRST = 3.6 V and VPIX = 3.3 V. The conversion gains were determined from the mean-variance photon transfer curve and the external electronic gain at 15.5 μV/e⁻ and 46.8 μV/e⁻ for the low and high gain readout paths, respectively.
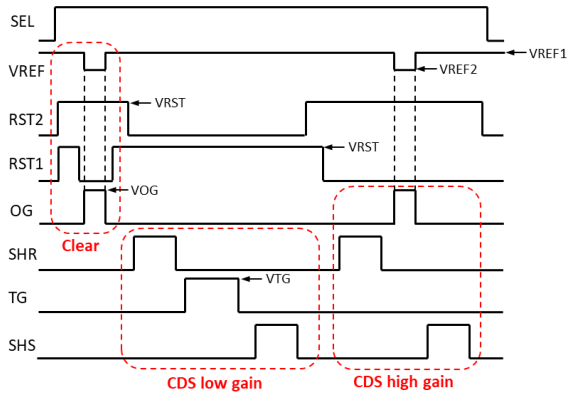
Fig. 4. Timing diagram used for pixel characterization showing the amplitudes of the control signals. The high levels of the signals not indicated are 3.3 V. All low levels are 0 V.
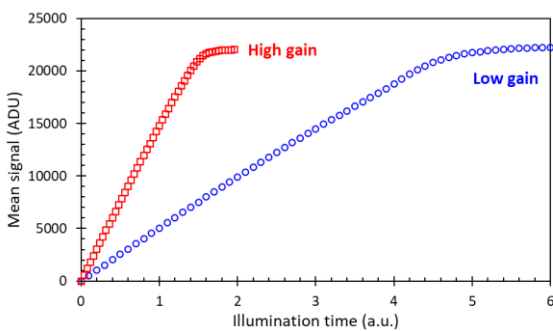

Fig. 5. Photoresponse at both gains. One ADU is 80.1 μV.

The two test images in Fig. 6 were taken under identical illumination and integration times. While the image at high gain in Fig. 6(a) is saturated, the image in Fig. 6(b), taken at low gain, has maximum signal of about a third of the FWC. The pixels having the buried channel implant had very similar characteristics, as can be seen in the right-hand side in Fig. 6.
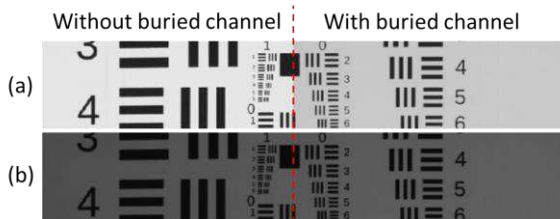

Fig. 6. Test pattern image at high gain (a) and at low gain (b) under the same illumination and readout conditions. The pixels in the left half of the image use surface channel transfer between the PPD and the sense node. The pixels in the right half have a buried channel as in Fig. 1.

The FWC and the readout noise for the low and high gain paths were measured to be 114 ke⁻ and 6.6 e⁻ RMS, and 38 ke⁻ and 2.4 e⁻ RMS, respectively. The gain ratio is very close to 3, and the dynamic range is 47500 (93.5 dB). The readout noise was measured without the shot noise from the dark current. This was accomplished by not pulsing the transfer gate and the OG in Fig. 4. With the dark current included, the noise at high gain rises to 3.6 e⁻ RMS, measured at 20 fps readout rate and 23 °C operating temperature.

In one of the reference 4T pixels the FWC is 22 ke⁻ for a CVF is 68 μV/e⁻, and despite the mean noise being lower at 1.63 e⁻ RMS due to the higher conversion gain, the DR is only 82.5 dB.

The design was characterized only in rolling shutter mode to demonstrate its operating principles, but it can also be operated in global shutter mode. Due to the signal storage under the sense gate, the readout path with high conversion gain accomplishes the same CDS and reset noise suppression as in rolling shutter mode, and should therefore exhibit the same readout noise. The reset noise in GS low gain mode is not eliminated, however this is not critical because at large signals the photon shot noise is expected to be dominant.

If not cleared before charge transfer out of the PPD, the dark current measured at the sense node would include the signal generated by interface traps under the SG because the silicon surface there is not pinned. The dark current in high gain mode as a function of VOG, shown in Fig. 7, indicates that VOG must be higher than 1.5 V for efficient clearing of the dark signal when VREF2 = 1.0 V. Under these conditions, the dark currents in low and high gain modes are nearly identical and are close to the values in the reference 4T pixels.
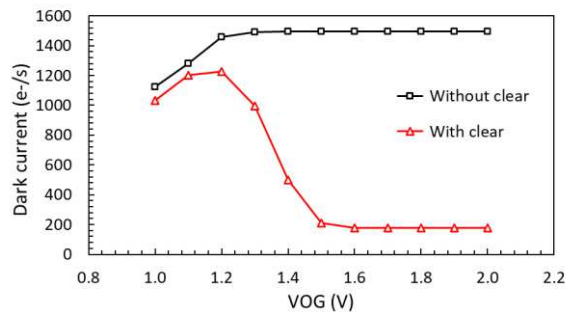

Fig. 7. Dark current in high gain mode with and without initial signal clear at VREF2 = 1.0 V and 25 °C.

Both the leading and the trailing image lag were measured because the initial dark signal clear could create a lag asymmetry. The leading edge lag was calculated as the normalized difference between the signals under steady-state illumination and in the first bright image following several dark images. Similarly, the trailing edge lag is the signal in the first dark image following sufficient number of bright ones, normalized to the signal under steady-state illumination. Experimentally it was established that a series of 5 bright and 5 dark images were more than enough to reach the steady state signal levels, as shown in the inset of Fig. 8.

The transfer out of the PPD was confirmed to be lag-free for VTG > 2.7 V when the TG pulse length is 1.5 μs. For the nominal VTG = 3.3 V the lag is negligible for TG pulses as short as 50 ns. The lag performance is very good because the transfer gate is relatively wide, and so is the SG, which behaves as a large charge collecting element.
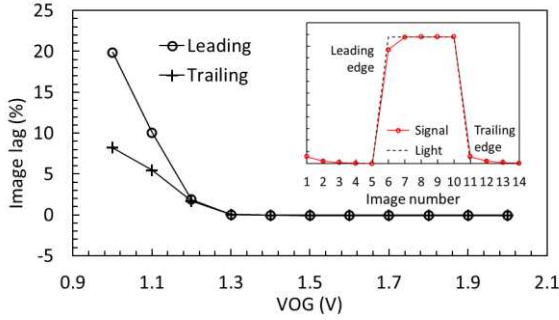
Fig. 8. Image lag as a function of VOG in high gain mode for the bias voltages listed in Section IV and signal at half FWC ($\approx$18 ke$^-$) in a pixel without buried channel.

The transfer from the SG to the SN has negligible lag for VOG $\geq$ 1.3 V as shown in Fig. 8 for pixels without the additional buried channel. At VOG = 1.5 V the low lag performance is maintained even for OG pulse lengths below 100 ns. The pixels implementing a buried channel have relatively low channel dopant concentration, and correspondingly low channel potential. Since the image lag for surface charge transfer pixels is already very low, they do not seem to offer any visible advantages.

V. CONCLUSION

In this work we present a new PPD-based pixel design using a floating gate to accomplish two consecutive signal readouts with different conversion gains. The pixel operates with a single exposure and a single charge transfer out of the PPD. The first prototype demonstrates significantly increased DR and negligible image lag. Although this implementation is for a pixel on 10 µm pitch, the proposed architecture could be scaled down to smaller pixels. Further improvements to the DR could be achieved by reducing the sense node capacitance and the readout noise, and also by increasing the effective sense gate capacitance. The proposed pixel architecture could be attractive for HDR imagers using single exposure.

REFERENCES

[1] I. Takayanagi and R. Kuroda, "HDR CMOS Image Sensors for Automotive Applications," IEEE Transactions on Electron Devices, vol. 69, no. 6, pp. 2815-2823 (2022).

[2] P. Vu, B. Fowler, S. Mims, C. Liu, J. Balicki, H. Do, W. Li and J. Appelbaum, "Low Noise High Dynamic Range 2.3Mpixel CMOS Image Sensor Capable of 100Hz Frame Rate at Full HD Resolution" in International Image Sensor Workshop, Hokkaido, Japan, (2011).

[3] T. Willassen, J. Solhusvik, R. Johansson, S. Yaghmai, H. Rhodes, S. Manabe, D. Mao, Z. Lin, D. Yang, O. Cellek, E. Webster, S. Ma and B. Zhang, "A 1280x1080 4.2µm Split-diode Pixel HDR Sensor in 110nm BSI CMOS Process," in International Image Sensor Workshop, Vaals, The Netherlands, (2015).

[4] C. Ma, Y. Liu, Y. Li, Q. Zhou, X. Wang and Y. Chang, "A 4-M Pixel High Dynamic Range, Low-Noise CMOS Image Sensor With Low-Power Counting ADC," IEEE Transactions on Electron Devices, vol. 64, no. 8, pp. 3199-3205 (2017).

[5] I. Takayanagi, N. Yoshimura, K. Mori, S. Matsuo, S. Tanaka, H. Abe, N. Yasuda, K. Ishikawa, S. Okura, S. Ohsawa, T. Otaka, "An Over 90 dB Intra-Scene Single-Exposure Dynamic Range CMOS Image Sensor Using a 3.0 µm Triple-Gain Pixel Fabricated in a Standard BSI Process," Sensors, vol. 18, no. 2, p. 203 (2018).

[6] N. Akahane, S. Sugawa, S. Adachi, K. Mori, T. Ishiuchi and K. Mizobuchi, "A sensitivity and linearity improvement of a 100-dB dynamic range CMOS image sensor using a lateral overflow integration capacitor," IEEE Journal of Solid-State Circuits, vol. 41, no. 4, pp. 851-858 (2006).

[7] Y. Fujihara, M. Murata, S. Nakayama, R. Kuroda and S. Sugawa, "An Over 120 dB Single Exposure Wide Dynamic Range CMOS Image Sensor With Two-Stage Lateral Overflow Integration Capacitor," IEEE Transactions on Electron Devices, vol. 68, no. 1, pp. 152-157 (2021).

[8] K. D. Stefanov, "Imaging Device", patent application US 2021/0217799 A1, EP3840366A1 (2021).

Ultra-sensitive CMOS image sensor capable of operating down to 200 ulx at 60 fps

Pierre Fereyre[1], Bruno Gili[1], Stéphane Gesset[1], Alexandre Charlet[2], Séverine André[1], Philippe Kuntz[1]
[1]Teledyne - e2v, Avenue de Rochepleine, BP123, Saint-Egrève, F-38521, France
[2]Teledyne - Anafocus, C. Isaac Newton, 4, 41092 Sevilla, Spain

Abstract— This paper presents a 10µm pixel pitch, front side illuminated (FSI), dual electronic rolling shutter (ERS) and global shutter (GS) image sensor with a fully depleted pinned photodiode (FDPD). At 950nm a quantum efficiency (QE) of 45% and a modulation transfer function (MTF) of 54%, Nyquist frequency, is demonstrated. The circuit is operated in a flexible way embedding low level digital processing for image quality improvement such as high dynamic range (HDR) and fixed pattern noise (FPN) correction.

Keywords- CMOS Image Sensor, near infrared, low-light, High Dynamic range, quantum efficiency, MTF

I   INTRODUCTION

The current technology trend for most applications is toward miniaturization of pixels compatible with high-resolution, large format CMOS image sensors. However, low-light applications require sufficient video rate and minimum signal-to-noise ratio (SNR) and spatial resolution to allow human interpretation of a complex scene. In some cases, image processing algorithms are used to enhance video while preserving details in the image [1]. Although the contribution of algorithms is effective, it is necessary that the basic information is as close to reality as possible. In low-light environments, for small pixels, ultra-low readout noise is required [2], while for larger pixels, photon shot noise quickly becomes the limiting factor in detector performance. With small pixels, the digital binning, equivalent to expanding the pixel size, augments the signal while the noise of each is added. For this reason, low light level applications require larger pixel sizes, very sensitive where the noise only contributes once: every photon matters. Therefore, the innovative FDPD was developed to maximise the signal-to-noise ratio and the Near-Infrared MTF at the same time. This work is a continuation of previous achievement [3] [4].

II   CHIP ARCHITECTURE

A.   Pixel structure

The pixel scheme is composed of a pinned photodiode and 5-transistors (5T) addressable by programmable signals allowing to operate either in RS or GS mode with exposure control. The RS mode is suitable for low light level vision requiring low readout noise. The internal sequencer also implements a GS mode with digital double sampling (DDS), alternating a reset frame and an image frame. The subtraction of the images is done off-chip. In this mode the "kTC" noise is minimized within the limit of the integration of the low frequency noise power. It was demonstrated that increasing the thickness of the detector as proposed with

FDPD will also improve the extinction ratio, confirming the benefit of using a thick silicon detector, for GS pixels [5].

B.   Analog-To-digital conversion

The circuitry includes a dual column-wise ADC that enables video rates exceeding 90fps in 12b-ERS and 50fps in 10b-GS (DDS). The ADC architecture is based on a double ramp that enables the conversion of higher pixel signal values. The differential voltage ramps are applied simultaneously on the reset and signal levels by capacitive coupling as represented Figure 1. This approach allows to cover a larger voltage dynamic at the input of the comparator.

C.   On-chip processing

The architecture includes low-level image processing that corrects residual fixed pattern noise with better than bit accuracy and offset suppression without loss of dynamic. The pixel matrix includes shielded columns and rows used to hold respectively the horizontal and vertical FPN value. The sensor embeds row memories for subtracting the FPN values within the data stream. In ERS, it is proposed to perform a high dynamic interframe exposure which is combined with a readout chain including a real time, pixel-wise tone mapping feature. A first short exposure is performed and then transferred to the floating node. Meantime, a second long exposure is integrated in the photodiode. The reading out sequence starts with the sampling of the signal value corresponding to the short exposure and then the reset. The difference is converted with a short ramp and determines when it exceeds a threshold equivalent to the dynamic range of the ADC divided by the ratio of exposures. The next step consists of a decision process of the transfer of the charges integrated during the long exposure, conditioned to the result of the previous comparison. If the threshold value is not exceeded, then the signal of long exposure overwrites the short one. In this case, the correlated double sampling (CDS) fully removes the reset noise and thus maximizes the signal-to-noise ratio in the weak signal range. Using the same approach, the FPN value corresponding to each of the long and short exposures are subtracted accordingly. The processing of the image is thereby made easier because the photon transfer curves are joined without gap, leading to an uncompressed and directly interpretable HDR display. The average temporal noise is reduced to the source follower and the ADC around 1.5 electron rms. The dynamic range with this design is proven at 110dB for an exposure ratio of 30dB.

III DEVICE STRUCTURE

A. Semiconductor process

The FDPD technology is based on 180nm foundry process and cost-effective compared to other approaches [6] [7] [8]. The FDPD principle is based on negative biasing of the bulk with non-zero electrical field in the full epi and the deep depletion enhancement (DDE) diffusion for preventing front-back current leakage (Figure 2). Therefore, the image is sharp and contrasted, the MTF being very close to the theoretical value in combination with a significantly improved QE. The pinned photodiode as well as the vertical n-well-bulk diodes are reverse biased and opposed to the vertical current flow. The p-well delimiting the pixels is connected to the ground. To oppose the direct conduction, the DDE is inserted between the p-wells and the bulk to form a floating NP junction polarized in such a way that it blocks the vertical current. The operation of the readout and processing circuitry is protected from the negative biasing of the bulk by using a deep n-well that creates a p-n junction. The TCAD simulation shown Figure 3 demonstrates that the depletion extends across the entire active silicon in the presence of the VBS static polarization as opposed to the same structure biased to ground. The deep depletion W is given by the following formula [9]:

$$W = \sqrt{\frac{2\varepsilon o \varepsilon i}{q N_D}\left(\frac{kT}{q}\ln\left(\frac{N_A N_D}{ni^2}\right) + V_{BS}\right)}$$

The two characteristics dimensioning the depletion depth are the doping or resistivity of the silicon and the V_{BS} voltage. For a depletion of some tens of microns and a voltage of some tens of volts, it is appropriate to consider a silicon doping of few 1E12 atoms per cm3, as depicted in Figure 4.

IV MEASUREMENT RESULTS

A. Quantum efficiency and MTF

The measurement of the quantum efficiency is performed on an optical bench comprising a monochromatic light source. The results are shown in Figure 5 in comparison with the theoretical calculation and a standard CMOS process implementation. The results are as expected. The gain in sensitivity is significant, especially in the near infra-red region, which is beneficial for the detection capacity in this spectral range, but also for the global SNR under low light level conditions. The MTF is measured using the slanting edge method and a tunnel with an optical aperture f/2.8. The value shown in Figure 6 is given for a spatial frequency of 50 lp/mm being the Nyquist limit corresponding to the pixel pitch. The effect of improving charge collection and reducing electron crosstalk is clearly visible as a function of polarization. In reference to the ground polarized silicon, this result shows a performance gain on the contrast essentially in the longest wavelengths, corresponding to the deepest penetration of the radiation.

B. Low light performance

The low light level performance is evaluated according to the NEI (noise equivalent irradiance) criterion, for a given exposure time and lens aperture. The NEI is calculated considering a color temperature of the light source of 2856K, an average lens transmission of 83% at f/1 and an exposure time of 1/60 second. Given the pixel characteristics, the corresponding integrated photo-response value is 3 e-/µlx-sec at the sensor plane which gives an equivalent NEI of 175 µlx at the scene level. The silicon based NEI at 950nm wavelength is 29 pW/cm² for exposure time 33ms, making this device among the best in its category with reference to other technologies [10]. Similarly, the irradiance corresponding to an SNR equivalent to 10 dB, which is the minimum for an acceptable image according to the ISO standard, is 1.2 mlx. This level corresponds to the starlight condition.

V CONCLUSION

The dedicated pixel technology, which includes a fully depleted photodiode in thick, ultra-high resistivity silicon, provides a signal-to-noise ratio sufficient to operate in deep darkness. The dynamic range of a single image reaches 110 dB within the scene, which allows the sensor to be flexible in many operating conditions. This circuit is the first of a new generation of CIS with enhanced performance in the near infrared. For future work, it is envisaged to combine it with a higher silicon thickness and a BSI. This opens the perspective of other applications such as medical imaging or intelligent transportation systems (ITS).

REFERENCES

[1] G. Zahi et S. Y. 2017, «Adaptive intensity transformation for preserving and recovering details in low light images,» *Computing Conference, London, UK,* pp. 262-271, 2017.

[2] J. Ma, S. Chan et E. R. Fossum, «Review of Quanta Image Sensors for Ultralow-Light Imaging,» *in IEEE Transactions on Electron Devices,* vol. 69 n°6, pp. 2824-2839, June 2022.

[3] P. Fereyre and al, "L2 CMOS image sensor for low light vision," *International Image Sensor Workshop, P20,* 2011.

[4] K. D. Stefanov, A. S. Clarke and A. D. Holland, "Fully Depleted Pinned Photodiode CMOS Image Sensor With Reverse Substrate Bias," *IEEE Electron Device Letters,* vol. 38 n°1, pp. 64-66, Jan. 2017.

[5] D. L. Stefan Lauxtermann, «Backside Illuminated CMOS Snapshot Shutter Imager on 50µm Thick High Resistivity Silicon,» *International Image Sensor Workshop, P32,* 2011.

[6] P. Jonghoon et al., «Pixel Technology for Improving IR Quantum Efficiency of Backsideilluminated CMOS Image Sensor,» *International Image Sensor workshop, R14,* 2019.

[7] M. U. Pralle, C. Vineis, C. Palsule, J. Jiang et J. E. Carey, «Ultra low light CMOS image sensors,» *SPIE Defense + Commercial Sensing,* vol. 11741 Infrared Technology and Applications XLVII, 28 April 2021.

[8] Oshiyama et al., «Near-infrared sensitivity enhancement of a back-illuminated complementary metal oxide semiconductor image sensor with a pyramid surface for diffraction structure,» *IEEE International Electron Devices Meeting (IEDM),* 2017.

[9] S. Sze, Physics of Semiconductor Devices, second ed., Murray Hill, New Jersey: Wiley, 1981, p. 77.

[10] J. -E. Communal, «Comparing camera sensitivity with Noise Equivalent Irradiance,» *6th Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS),* pp. 1-4, 2014.
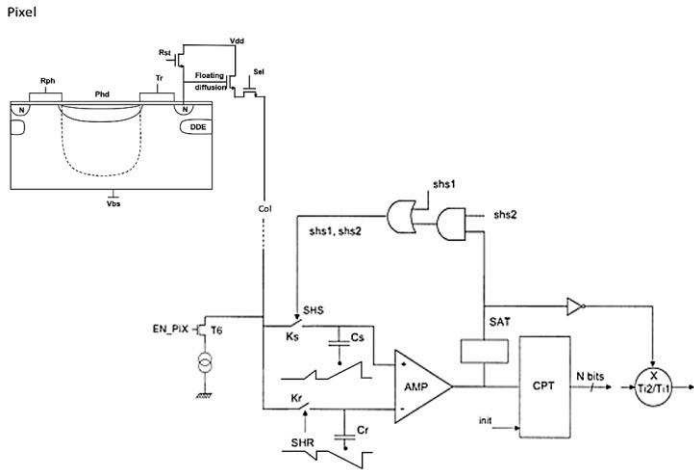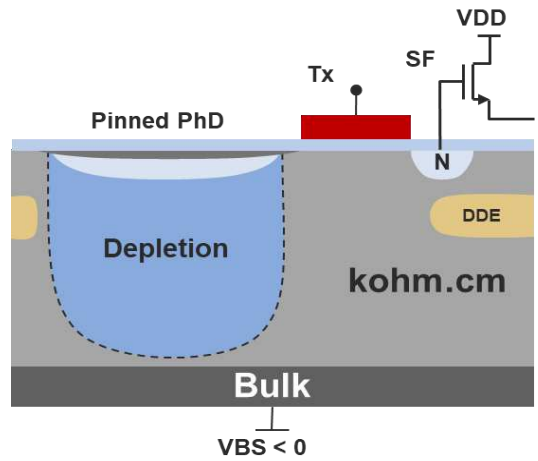
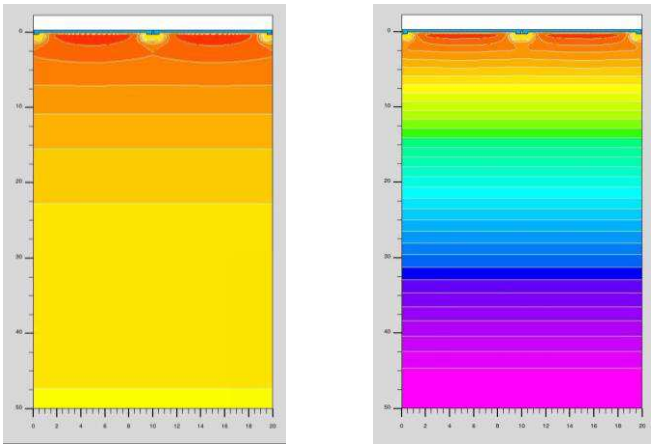**Figure 1 – HDR Pixel processing chain**

Figure 2 – Pixel structure

Figure 3 – FDPD equipotential, on the left grounded $V_{BS} = 0V$, on the right grounded $V_{BS} < 0V$
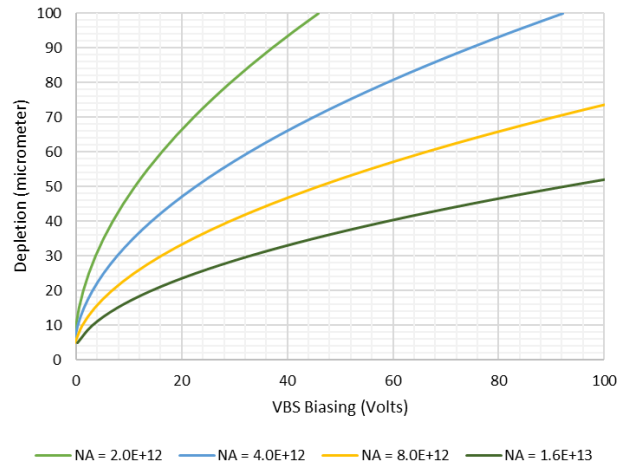

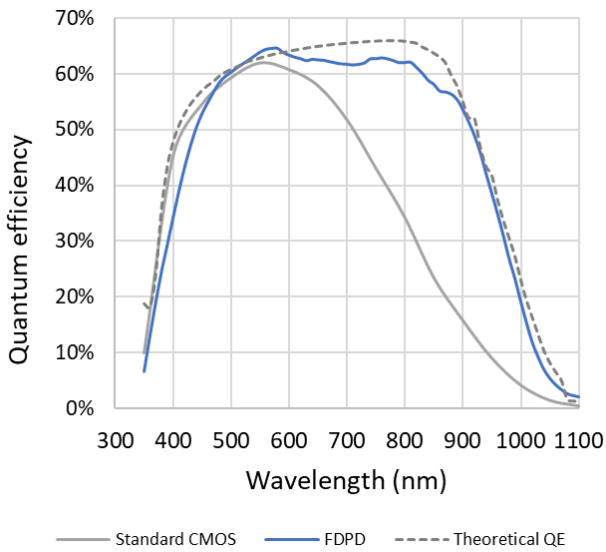**Figure 4 – Depletion depth factors**
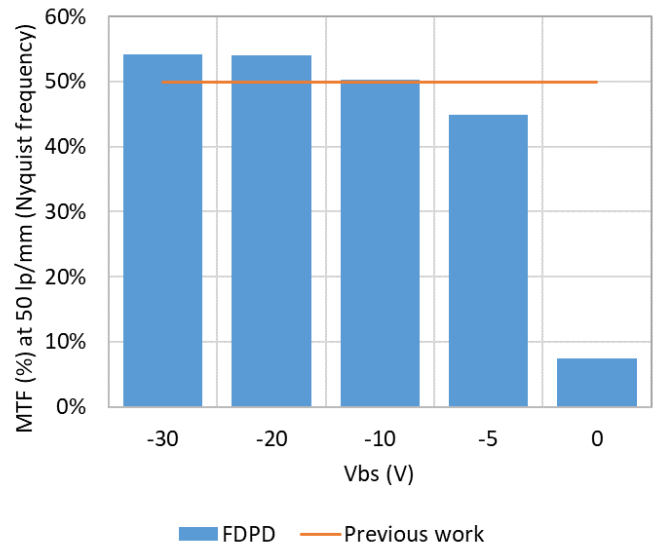

Figure 5 – FDPD Quantum efficiency

Figure 6 – MTF at 950nm

Evolution of a 4.6 μm, 512×512, ultra-low power stacked digital pixel sensor for performance and power efficiency improvement

Rimon Ikeno, Kazuya Mori, Masayuki Uno, Ken Miyauchi, Toshiyuki Isozaki, Hirofumi Abe, Masato Nagamatsu, Isao Takayanagi, Junichi Nakamura, Shou-Gwo Wuu[†], Lyle Bainbridge[‡], Andrew Berkovich[‡], Song Chen[‡], Ramakrishna Chilukuri[‡], Wei Gao[‡], Tsung-Hsun Tsai[‡], and Chiao Liu[‡],

Brillnics Japan Inc., Tokyo, Japan, [†]Brillnics Inc., Hsinchu, Taiwan, [‡]Reality Labs, Meta Platforms Inc., Redmond, WA, USA

Rimon Ikeno – tel: +81-3-6404-8801, e-mail: ikeno.rimon@brillnics.com

Abstract – We report improvement of a global shutter, stacked digital pixel sensor with 512 × 512, 4.6 μm pixels featuring an overlapped triple quantization scheme. It achieves an ultra-high dynamic range of 127 dB with reduced temporal noise and fixed pattern noise by pixel-design tuning and layout optimization. The new sensor chip achieves low power consumption of 5.8 mW, which is comparable to the original chip by design and operation optimizations despite the newly integrated voltage regulators for pixel power supply and pixel-control signals in the same die size as the original chip.

I. INTRODUCTION

Augmented Reality (AR) and Virtual Reality (VR) devices are emerging to be the next mobile computing platform. To meet the stringent performance, power and form factor requirements for AR/VR consumer devices, image sensors must be optimized for computer vision algorithms, which require global shutter (GS) operation, high sensitivity, high dynamic range (HDR), and ultra-low power consumption [1].

We have reported a GS digital pixel sensor (DPS) fabricated by a stacked process with pixel-level Cu-to-Cu hybrid-bonding (HB) interconnects between the two stacked layers [2, 3]. The sensor had a 512 × 512 pixel array with 4.6 μm DPS pixel and featured an overlapped triple quantization (3Q) scheme that performs a time-to-saturation quantization and the dual conversion gain (CG) linear ADC modes sequentially in the same frame to extend DR with a 10-bit ADC. It achieved an ultra HDR of 127 dB and low power consumption of 5.8 mW at 30 frames per second, which demonstrated the best figure of merit (FOM) among recently emerged 3D-stacked DPSs [4]-[8].

In this article, we report design and evaluation results of an improved version of the previously reported DPS. We made further circuit and process optimizations and achieved better performance than the original chip for productization. The new chip integrates on-chip voltage regulators into the 4mm × 4mm die, maintaining the same footprint as the original chip. These evolutions have made the DPS chip most suitable for battery-powered and always-on mobile computer vision applications.

II. SENSOR DESIGN AND OPERATION

Fig. 1 shows a circuit diagram and a cross-section of the stacked DPS pixel. The pixel is partitioned into two parts; a dual-CG type pixel with a backside-illuminated (BSI) pinned photodiode (PPD) in the CIS layer on the top, and an in-pixel ADC circuit with 10-bit SRAM in the ADC layer at the bottom. These two layers are connected using HB technology [9, 10] in each pixel. In Fig. 2, a detailed circuit diagram of the stacked DPS pixel illustrates the in-pixel ADC circuit and 10-bit SRAM with the logic circuit in-between them to enable the 3Q quantization scheme described hereafter.

The DPS features an overlapped 3Q scheme that performs time-to-saturation (TTS mode) quantization for high-light signal, high-CG linear ADC (so-called PD-ADC mode) for low-light signal, and low-CG linear ADC (so-called FD-ADC mode) for middle-light signal sequentially in one frame [1, 2]. The in-pixel ADC circuit automatically selects the appropriate quantization mode based on the received light of each pixel and stores the quantized value in the pixel memory. Pixel-signal timing of overlapped 3Q scheme is illustrated in Fig. 3. A typical photo-response curve of the overlapped 3Q DPS is illustrated in Fig. 4.

In the new sensor design, we made further

optimization of device sizes in the pixel and modified metal-wire layout throughout the pixel array. These changes reduce temporal noise (TN) and fixed pattern noise (FPN) by lowering and balancing the coupling capacitance between sensitive nodes.

The chip block diagram in Fig. 5 shows circuit components on the CIS and ADC layers with their HB connections. The new sensor chip integrates charge pumps to generate a higher voltage than the analog supply (2.5V) and a negative voltage lower than the ground. The charge-pump outputs and the primary voltage supplies drive on-chip low-drop-out (LDO) regulators that supply the pixel-array in the ADC layer and the pixel-signal drivers in the ADC and CIS layers. Despite the additional components for internal voltage regulation, the new chip was laid out on the same die size as the original chip (4mm × 4mm). Fig. 6 is a photomicrograph of the stacked chip in a chip-scale package (CSP).

III. CHARACTERIZATION RESULTS

Fig. 7 shows the pixel-signal histogram at the dark condition for the original and new chips. The signal distribution of the new chip is narrower than the original chip. This improvement is due to the reduced pixel-wise FPN resulting from the metal-wire layout optimization in the pixel array for coupling-capacitance reduction and balancing. SNR drop at the junction point of the high CG and low CG ADC modes is improved by tuning the DCG capacitor [11], while maintaining the 127-dB DR.

In Fig. 8, power consumption of the original and new chips is compared with different integration times (Tint). Although the on-chip voltage regulators are integrated in the new chip, its power is almost the same as that of the original chip in the 1-ms Tint case. In the longer Tint cases, the new chip consumes less power than the original chip. The lower power consumption of the new chip is due to circuit-design optimizations of the peripheral analog modules and improved PLL control which reduces the PHY power consumption.

Fig. 9 shows an image captured by the 3Q scheme.

Table 1 compares the sensor performance index for recent stacked pixel- or cluster-wise ADC sensors. In comparison with the original chip in the previous work, the new chip has smaller noise floor (TN) and FPN as expected by the pixel-design improvement discussed in this article. As a result, the new chip has better FOM than the original chip, which has superior FOM than the other references in the table.

IV. SUMMARY

We have developed the second-generation chip of a stacked digital pixel sensor with an overlapped triple quantization scheme. It integrates voltage regulators for pixel power supply and pixel-control signal drivers in the same die size as the original chip. The new chip has improved temporal noise and fixed pattern noise performance achieved by pixel-design tuning and layout optimization. The sensor realizes the best FOM among the recent stacked pixel- or cluster-wise ADC sensors.

ACKNOWLEGMENT

REFERENCES

[1] C. Liu, et al., "Intelligent Vision Systems – Bringing Human-Machine Interface to AR/VR", in *IEDM Tech. Dig.,* San Francisco, CA, USA, pp.218-221, 2019.

[2] C. Liu et al., "A 4.6 μm, 512×512, ultra-low power stacked digital pixel sensor with triple quantization and 127 dB dynamic range," in *IEDM Tech. Dig.*, San Francisco, CA, USA, Dec. 2020, pp. 327–330.

[3] R. Ikeno et al., "A 4.6-μm, 127-dB Dynamic Range, Ultra-Low Power Stacked Digital Pixel Sensor With Overlapped Triple Quantization," *IEEE Trans. Electron Devices*, vol. 65, no. 6, pp. 2943-2950, Jun. 2022.

[4] K. Mori et al., "A 4.0 μm Stacked Digital Pixel Sensor Operating in a Dual Quantization Mode for High Dynamic Range," *IEEE Trans. Electron Devices*, vol. 65, no. 69, pp. 2957-2964, Jun. 2022.

[5] M. W. Seo et al., "A 2.6 e-rms low-random-noise, 116.2 mW low-power 2-Mp global shutter CMOS image sensor with pixel-level ADC and in-pixel memory," in *Proc. Symposium. VLSI Tech.*, Jun. 2021, pp. 1–2.

[6] M. Sakakibara et al., "A 6.9-μm pixel-pitch back-illuminated global shutter CMOS image sensor with pixel-parallel 14-bit subthreshold ADC," *IEEE J. Solid-State Circuits*, vol. 53, no. 11, pp. 3017–3025, Nov. 2018.

[7] T. Takahashi et al., "A stacked CMOS image sensor with array-parallel ADC architecture," *IEEE J. Solid-State Circuits*, vol. 53, no. 4, pp. 1061–1070, Apr. 2018.

[8] H. Sugo et al., "A dead-time free global shutter CMOS image sensor with in-pixel LOFIC and ADC using pixel-wise connections,"

in *Proc. Symp. VLSI Circuits*, Jun. 2016, pp. 1–2.

[9] C.-T. Ko, et al., "Wafer-level bonding/stacking technology for 3D integration," *Microelectron. Rel.*, vol. 50, no. 4, pp. 481–488, Apr. 2010.

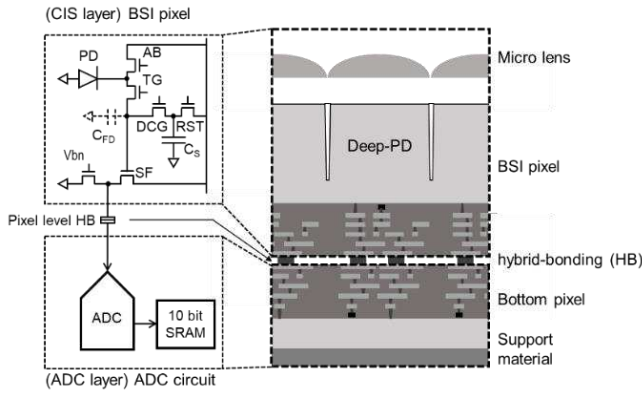[10] P. Ramm, et al., *Handbook of Wafer Bonding*. Hoboken, NJ, USA: Wiley, 2012.

[11] N. Akahane, et al., "Optimum Design of Conversion Gain and Full Well Capacity in CMOS Image Sensor With Lateral Overflow Integration Capacitor," *IEEE Trans. Electron Devices*, vol. 56, no. 11, pp. 2429-2435, Nov.. 2009.

Fig. 1. Circuit/Block diagram and cross sectional view of the stacked digital pixel sensor.
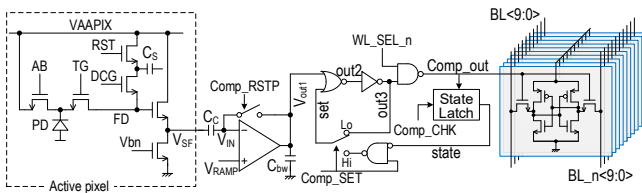
Fig. 2. Detailed circuit digram of the stacked DPS pixel.
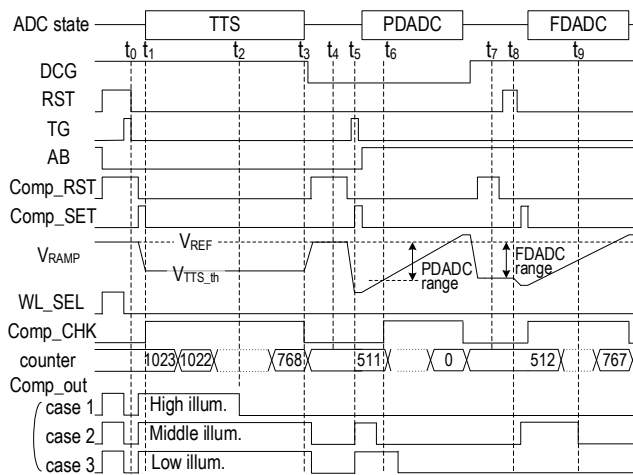
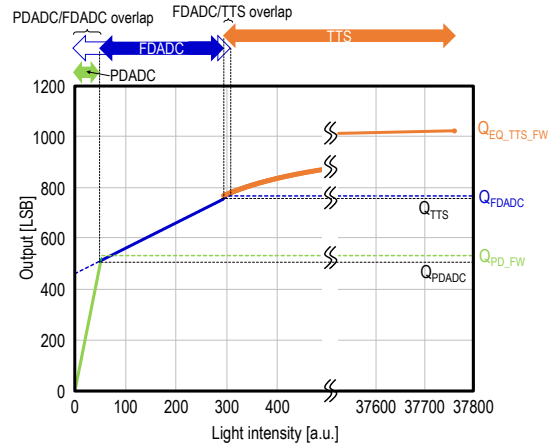Fig. 3. Timing diagram of overlapped 3Q operation.

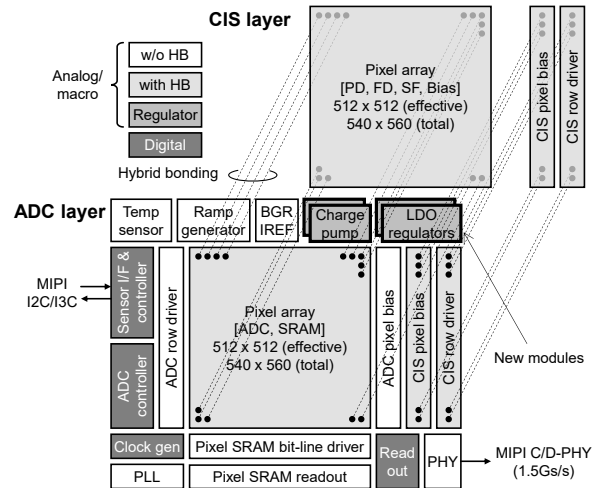Fig. 4. Photo-response curve of the overlapped 3Q DPS.
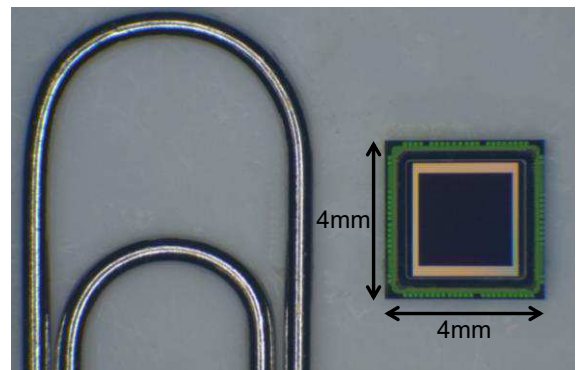

Fig. 5. Sensor chip block diagram.

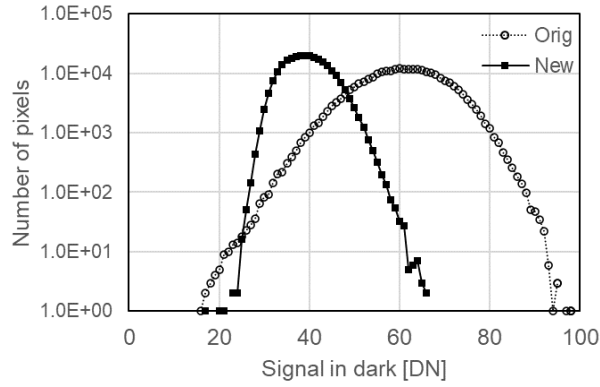Fig. 6. Sensor chip photomicrograph with a paper clip.

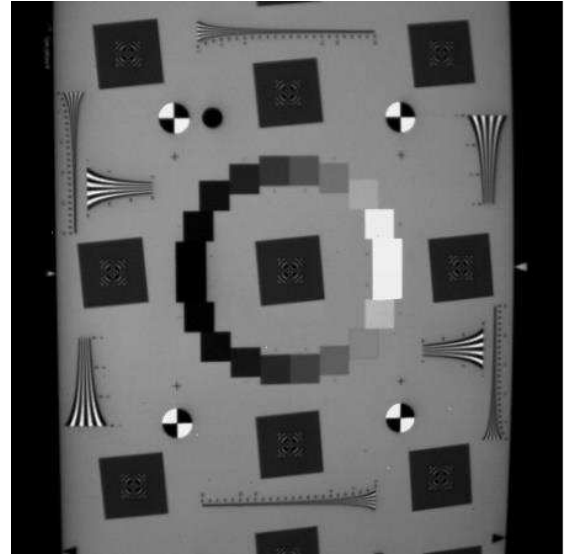Fig. 7. Dark histogram showing the FPN improvement in the new chip in this work.
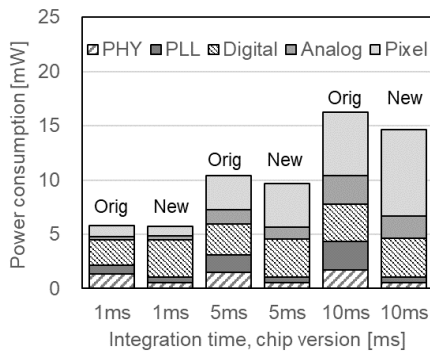

Fig 9. A test chart image captured using 3Q scheme.

Fig. 8. Power consumption of the original and new chips. at different integration times

Table. 1 Sensor performance matrix for recent stacked pixel wise ADC system.

Specification	This work	Previous work [2,3]	VLSI2021 [5]	ISSCC 2018 [6]	JSSC2018 [7]	VLSI2016 [8]
Process technology	45nm/65nm	45nm/65nm	65nm/28nm	90nm/65nm	90nm/55nm	45nm/65nm
Pixel size [µm]	4.6	4.6	4.95	6.9	4.8	1.65
# of pixels (H×V)	$512^H \times 512^V$	$512^H \times 512^V$	$1668^H \times 1364^V$	$1632^H \times 896^V$	$2360^H \times 1728^V$	$2576^H \times 1920^V$
In-pixel memory bit #	10b (1 pixel/ADC)	10b (1 pixel/ADC)	22b (1 pixel/ADC)	14b (1 pixel/ADC)	12b (160 pixels/ADC)	12b (16 pixels/ADC)
QE (@530nm) max [%]	96 (Mono)	96 (Mono)	NA	NA	N/A	NA
Dynamic range [dB]	127	127	74[1]	70.2	69	NA
Conversion gain [µV/e⁻]	170/12	170/7	132	NA	65	NA
Linear full well [ke⁻]	5/34/9000[2]	3.8/51/9000[2]	14	16.6	6.8	220
Noise floor [e⁻]	4.0	4.2	2.6 (24dB gain)	5.15	2.4 (24dB gain)	NA
Dark FPN [e⁻]	27	47	1.94/0.45	NA	N/A	NA
Power [mW]	5.8	5.8	497.8	746	1340	NA
FOM[3]	0.0608[4] 0.0013[5]	0.1809[4] 0.0014[5]	1.74 (24dB gain)	1.24	15.84 (24dB gain)	NA

(Notes)   (1) Estimation / (2) Equivalent FWC estimated with photo response plot

(3) Figure of Merit (FOM) is based on the following formula [5]; $FOM = \frac{(power) \times (noise)}{(\# \text{ of pixels}) \times (frame\ rate) \times (DRU)}$, $DRU = \frac{(saturation)/(gain)}{(noise)}$

(4) Without FPN correction / (5) With FPN correction