320x240 SPAD direct Time-of-Flight Image Sensor and Camera based on In-Pixel Correlation and Switched-Capacitor Averaging

Maarten Kuijk[1] (mkuijk@vub.be), Ayman Morsy[1], Thomas Lapauw[1], Thomas Van den Dries[1], Mohamed.A.Bounouar[2], Hans Ingelberts[1], Daniel Van Nieuwenhove[2].
[1]ETRO.RDI, Vrije Universiteit Brussel, Pleinlaan 2, 1050 Brussel, Belgium,
[2]SONY Depthsensing Solutions, Rue Jules Cockx 8, 1160 Brussel, Belgium.

Correlation-Assisted Direct Time-of-Flight (CA-dToF) is demonstrated for the first time with a large SPAD array. SPAD triggers are processed in-pixel, avoiding information congestion. This is accomplished by sampling two orthogonal triangle waves, correlated with a VCSEL floodlight source emitting 2-ns pulses to the scene. Using small switched-capacitor circuits, exponential-moving-averaging (EMA) is applied on the two streams of sampled voltages, delivering two analog voltages (VQ2, VI2) defining a phase shift indicating the ToF distance. Ambient photons and dark counts are averaged out, their shot noise remains. The camera captures depth-sense images with a promising sub-cm precision over a 6-m range of detection even with a PDE as low as 0.7%.

CA-dToF is presented on a 32×32 prototype array in [1], based on sine and cosine demodulation using single stage EMA and in [2], with 100-klux solar ambient light suppression.

We now demonstrate the operation of a large QVGA sensor with pixels that make use of *two-stage cascading*: the first stage samples and averages at each SPAD's trigger occurrence, the second stage samples and averages the output of the first stage at a fixed clock rate, thereby reaching a total averaging capacity of $200 \times 400 = 8\times10^4$. For ease of on-chip generation, triangular (TSIN, TCOS) instead of sinewave signals are this time applied as correlation inputs

Classical direct and indirect ToF sensors *time-bin* the time-of-arrivals (ToAs) during acquisition, thereby losing *precious timing* information and thus deteriorating the distance precision. With the presented correlation and averaging approach, no time-discretizing is applied, which besides improving the precision, will also allow to easily locate the laser photons' center of mass, thereby *effectively averaging out SPAD-timing jitters, system jitters and uncorrelated ambient light inputs.* The latter includes light pulsation from fluorescence tubes and LED lighting.

Figure 1 shows the camera circuit with the image sensor chip fabricated in a high-voltage 180-nm CMOS technology from X-FAB having a qualified SPAD. No 3D stacking, or BSI technology has been used.

Little communication is needed between the image sensor and the PC, because the two analog voltages (VQ2, VI2) defining the ToF distance, are computed *in-pixel.* At the end of each sub-frame, these values are read out through on-chip column ADCs that have settable resolution (8b...11b) for evaluating the resolution requirements.

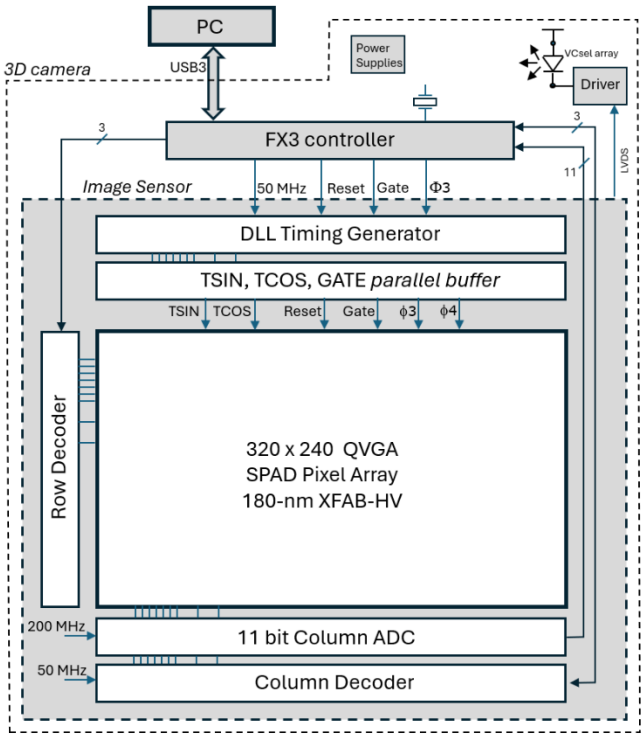
An FX3 microcontroller (Infineon) provides this communication with the PC over USB3 and also controls the direct memory access with the image sensor chip. The sensor drives the laser driver chip (EPC21603) by low-voltage differential signaling (LVDS), which on its turn drives the VCSEL array (EGA2000-850-N from OSRAM-AMS) for generation of strong short 2-ns light pulses in sync with the triangle waves.

The orthogonal triangular waves are generated on-chip in sync with the clock provided by the FX3. A DLL timing generator divides a 40-ns period into a grid of 64 points with 625-ps spacing, for defining all time-dependent operations including the command of the laser and the triangular corner points: this allows the system to flexibly switch between 25 MHz and 100 MHz triangular waves in-between subframes.

Figure 2 shows the pixel circuit having an area of 30×30 μm^2 having one SPAD (Ø=10-μm) delivering a fill factor of 8.7%. At chip-submission, AC-coupling was the only way to make arrays, and quite a large area for C_0 was needed to ensure a reliable connection between SPAD and circuit. Additional in-pixel circuits were integrated (not shown) including an array-wide time-gate, settable on the 625-ps grid. At the 850-nm wavelength and 2.4V excess bias, the PDP will be ~ 8%, resulting in an overall (low) PDE of ~ 0.7%. Besides the two averaging stages for the two correlations, there is also a photon-counter for imaging the light intensity (with and without laser illumination), the exposure time being determined by the FX3. For calibrating the photon counter's gain, we made use of the statistics of the photon shot noise.

Figure 3 shows how the two EMA stages reduce the noise elegantly after each step. Note that already early in the process a first good estimation is found, and that saturation is not possible. The number of ambient photons (**A**) and the number of ToF signal photons (**S**) are indicated. Dark current photons can be categorized into class A because they arrive at random uncorrelated moments.

In the demonstrations, 20-ms subframes (at 24 fps), are used. This length is quite long but is reasonable in view of the very low PDE. A 40-nm optical bandpass filter at 850 nm wavelength is used to reflect most of the ambient light photons. **Figure 4** demonstrates the operation at 25 MHz, giving a 6-m unambiguous range. Two subframes are recorded (0°, 180°) *cancelling the offsets in the signaling chain,* like mismatches from the voltage follower's transistors and the ADC's offsets. Besides the grey and distance estimate images, the differential Q2 and I2 images are also shown (none of which have been averaged in the PC). A cross-section (row 124) gives more quantitative values. The instantaneous and average depth, and the differential Q2 and I2 values are shown.



← **Figure 1:** The camera is controlled by a PC over USB3. An FX3 microcontroller provides the data-handling. The 50 MHz interface clock from the FX3 controller is used at several places inside the image sensor chip.

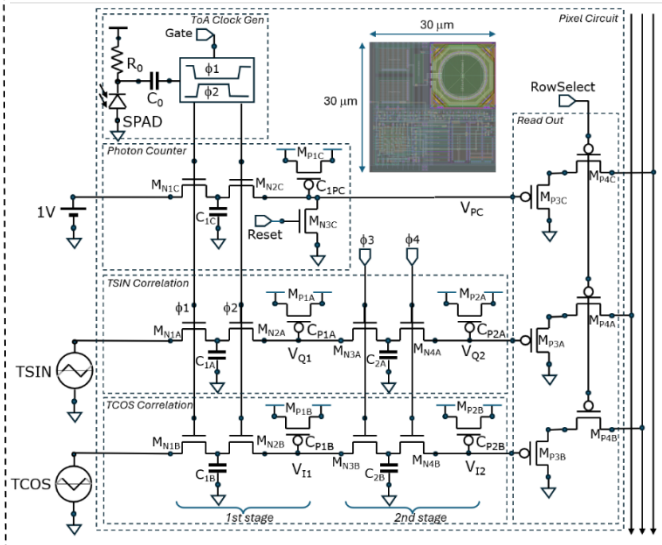


Figure 2 ↑: The pixel circuit: each ToA is converted to non-overlapping clocks (Φ1, Φ2) for toggling the photon counter stage, the TSIN and TCOS correlation's samplers and first averaging stages. Read out is through PMOS voltage followers. ↑

Figure 3 ↓: Statistical simulation. V_{Q1} and V_{I1} are averaged signals from the input that contain: random values between 0.1 and 0.9V in response to ambient triggers, and the values from the signal photons at 18 ns. They subsequently serve as input for getting averaged by the second stages leading to voltages V_{Q2} and V_{I2}, from which the ToF delay is estimated.

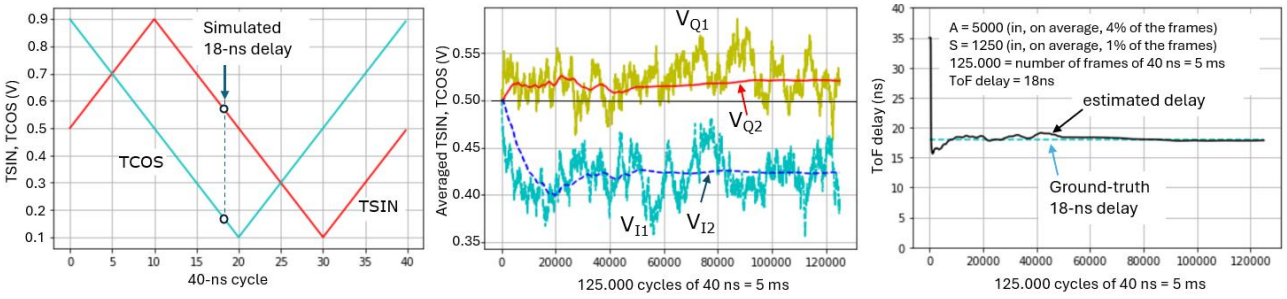


Figure 4: Demodulation using (0˚, 180˚) phases at 25 MHz: a = mannequin, b = whiteboard, c = reflectors for testing dynamic range, d = closet in corridor, e = box with black & white stripes for testing color dependency. ↓

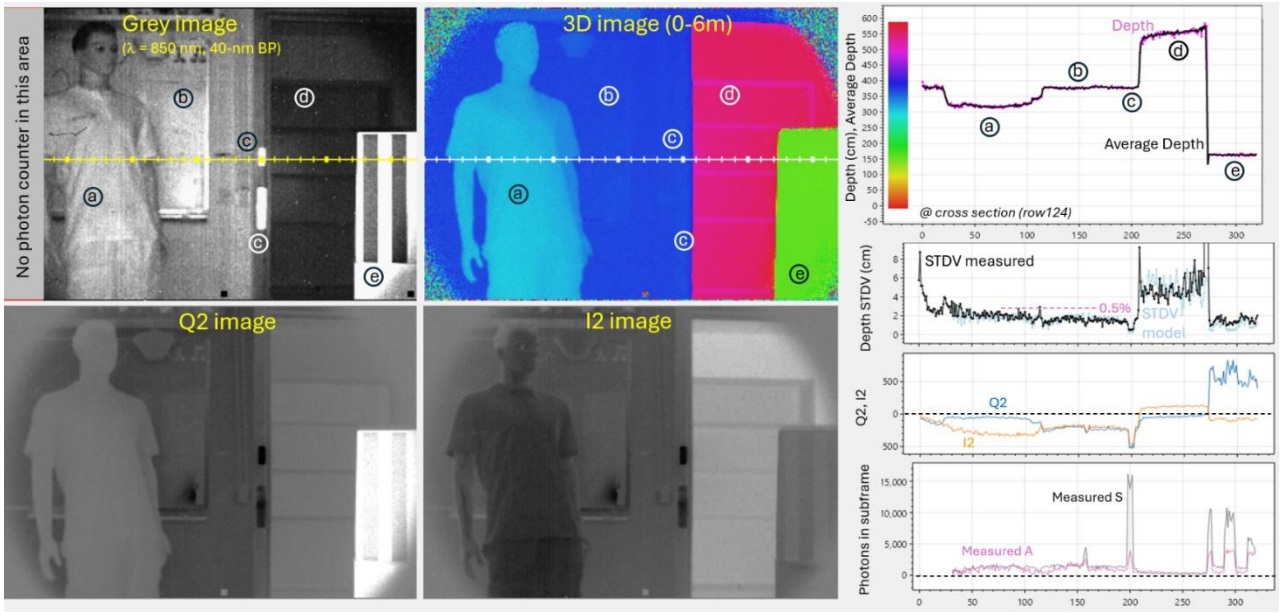


Figure 5: Demodulation at (0˚, 180˚) phases at 25 MHz and (0˚, 180˚, 90˚, 270˚) phases at 100 MHz. Ambient is 1 klux at the whiteboard (b) and 2 klux at the box (e). ↓

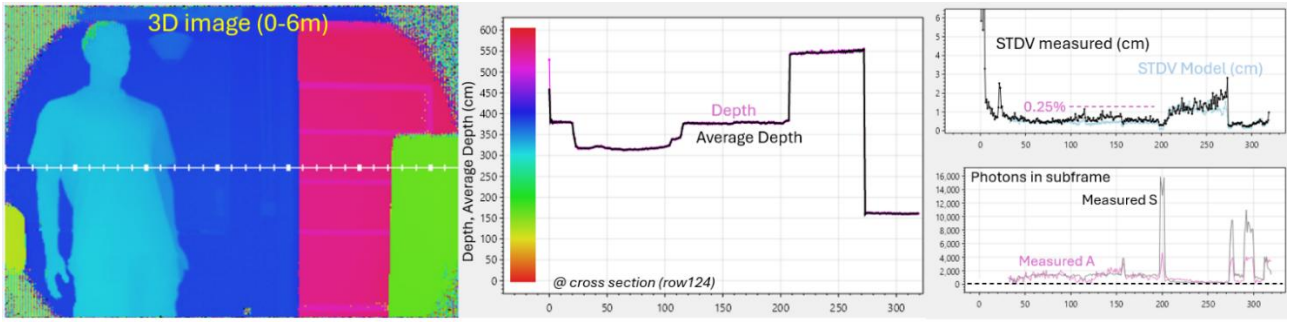


Figure 6: Same conditions as Fig. 5 but now including a 50% nearest neighbor spatial filter. Fixed pattern distance noise (in cm) is demonstrated on the "flat" surface of box (e). ↓

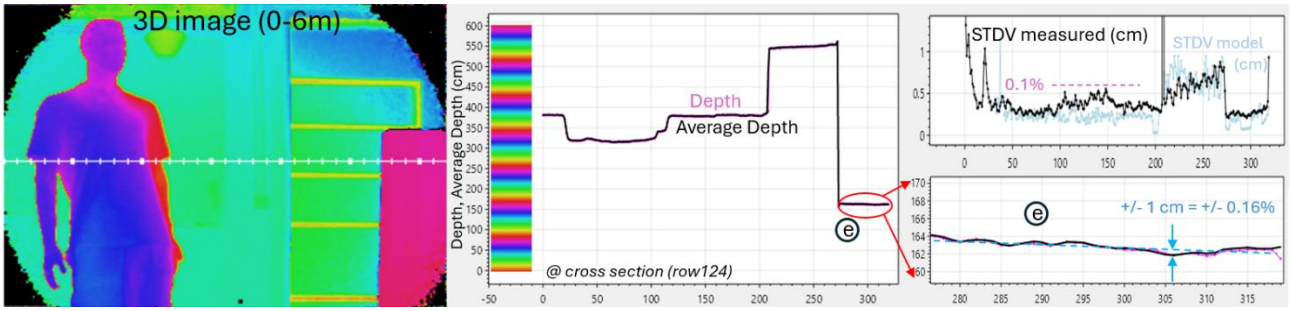
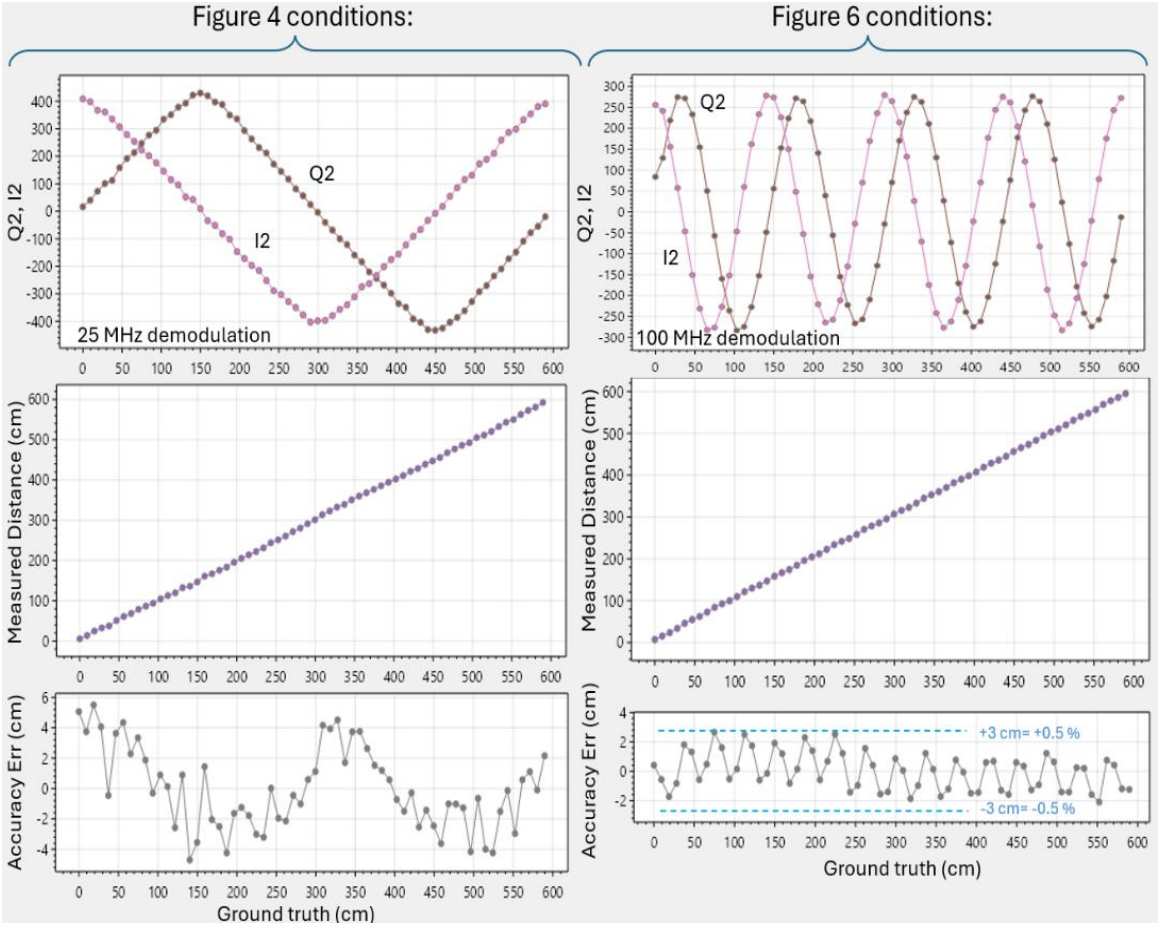


Figure 7: Accuracy on a pixel in the center of the image. The laser is cycled over the full 360˚ phase range in 64 steps. The level of cyclic error is then obtained by plotting the accuracy error being the difference between the calculated distance (derived from Q2 and I2) and the groundtruth. It can be concluded that the 100 MHz demodulation also helps accuracy (improving from +/-1 % to +/- 0.5%). A shorter laser pulse helps to reduce this cyclic error even further (not shown). ↓



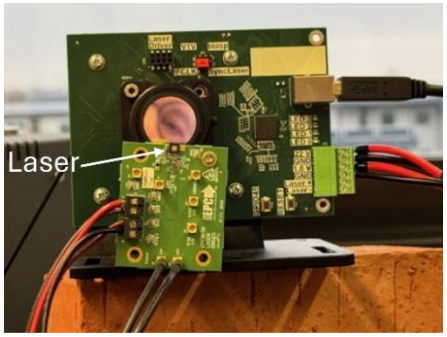


Figure 9: With a moveable whitewall, measurements were performed with a calibrated solar simulator device (Hal-302). Estimated distance and STDV are shown under 4 ambient light conditions →

← **Figure 8**: Picture of the camera without external power supplies.

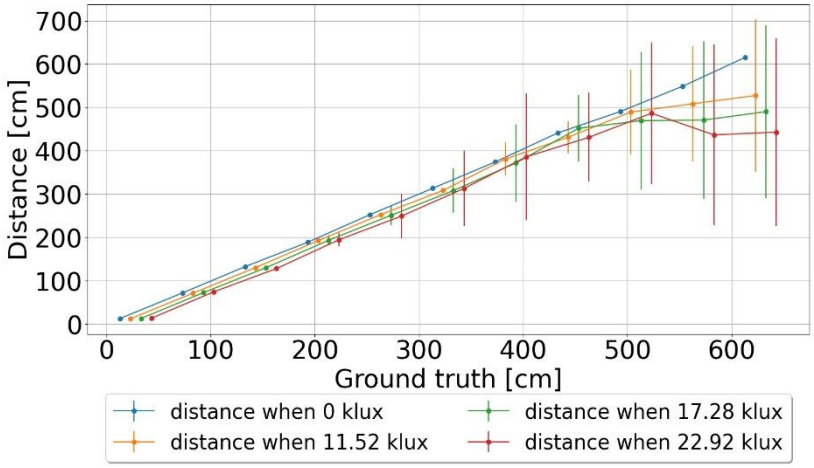


Table 1: Summarizing performance ↓

	Image Sensor supply		SPAD		Performance				System	
conditions of	Supply Voltage	Image Sensor Current	Supply Voltage	Current	Sub-framerate	3D image framerate	Accuracy over 360° phase	STDV (@3.5m, 1 klux, 50% reflectivity)	ADC	Interface Clock
unit	(V)	(mA)	(V)	(mA)	(fps)	(fps)	+/- %	(%)	(bits)	(MHz)
figure 4	1.8V	61 mA	22 (2.4 excess)	1 mA (at 1 klux)	24	12	1	< 0.5	10	50
figure 5		98 mA			24	4	0.7	< 0.25		
figure 6		98 mA			24	4	0.5	< 0.1		

Also shown are Q2 and I2, the number of ambient (A) and signal (S) photons per 20-ms subframe, extrapolated from photon counter measurement frames with and without the laser being turned-on. This is useful for comparing in real-time the standard deviation (STDV) model with the measured STDV, also shown.

Four subframe phases (0°, 180°, 90° & 270°) offer *non-linearity* suppression: as with iToF, this improves the accuracy of the estimated distance. To greatly improve the STDV, modulation at a higher frequency will help: it improves in proportion to the frequency increase. To achieve both, i.e. high unambiguous range (6m) and good precision, a rough distance estimate can first be made with 2 subframes at 25 MHz, followed by a much more precise and accurate *4-phased* estimate at *100 MHz* (**Fig. 5**). From the latter, four distances in the 6-m range become available (on a 1.5-m grid), and in real-time the one that is closest to the 25-MHz estimate, can then be selected. A similar approach in an iToF system was followed in [3]

Additionally, in **Fig. 6**, a nearest neighbors' spatial filtering finally gives sub-cm precision. This allows the color scale to be cycled 7 times over the 6-m range, the visible low noise in the depth image then shows that there is also low noise in the depth itself. The *predicted* depth standard deviation (STDV model) is indicated in light blue based on the measured A and S and aligns nicely with the *measured* STDV. Furthermore, without any distance calibration performed, an example of the fixed pattern distance noise is given at position **e**, where a depth variation of about 1 cm is achieved at the 164 cm distance (+/- 0.16 % for 6-m unambiguous range).

Figure 7 shows accuracy (low cyclic jitter), **Fig. 8** is a picture of the camera, and Table 1 summarizes the system's performance.

Figure 9 shows a measurement with higher ambient light levels. Having only a low PDE, two VCSEL arrays of the same type were used for illumination of the scene emitting with in total 520 mW of average light power. Fig. 6 camera settings were used, however this time *a third averaging stage* (not shown in Fig. 2) was unlocked in each pixel to increase the averaging capacity to 200 × 200 × 200 = 8 × 10^6 samples. In this way, low to very high ambient light levels can be accommodated simultaneously on different pixels in a same frame. Lower light levels will just be oversampled without harm for the precision and accuracy. The clock periods for the 2nd and 3rd stage were chosen at 1 µs and 100 µs respectively, in this way achieving a geometric spreading which allows the three stage EMA system to operate optimally. The estimated distances are shown in a skewed way to better distinguish between the 4 ambient light conditions (0, 11.5, 17.3 and 22.9 klux). Up to 300 cm, acceptable precision and accuracy is present. Above that distance, the precision worsens further to levels where the method of choosing the correct distance (out of 4), starts to fail, impacting accuracy.

Conclusion: CA-dToF and the use of switched capacitors in image sensors prove to be promising for depth-imaging applications even in large arrays demonstrated by the achieved good precision, accuracy and low power merits. Using a BSI and/or 3D stacking SPAD image sensor technology, a PDE that is one to two orders of magnitude better would be in reach, allowing to improve the CA-dToF performance accordingly.

Acknowledgements: This work was partly funded by SRP78 and CoDaFlight.

[1] A. Morsy, M. Kuijk. *Sensors*. 2024; 24(16):5380. https://doi.org/10.3390/s24165380

[2] A. Morsy, C. Baijot, et al, *IEEE Sensors Letters*, vol. 8, no. 10, pp. 1-4, Oct. 2024, doi: 10.1109/LSENS.2024.3453038

[3] C. S. Bamji et al., IEEE Journal of Solid-State Circuits, vol. 50, no. 1, pp. 303-319, 2015, doi: 10.1109/JSSC.2014.2364270.