A 2-Tap Macro-Pixel-Based Indirect ToF CMOS Image Sensor for Multi-Frequency Demodulation

Peyman F. Shahandashti*, P. López* V.M. Brea*, D. García-Lesta*, Miguel Heredia-Conde†
*Centro Singular de Investigación en Tecnoloxías Intelixentes (CiTIUS), Universidade de Santiago de Compostela, Spain
†Center for Sensor Systems (ZESS), University of Siegen, Germany
E-mail: peyman.fayyaz@usc.es, p.lopez@usc.es, heredia@zess.uni-siegen.de

***Abstract*—Indirect time of flight (IToF) allows for accurately retrieving 3D geometry without the need for exorbitant time resolution. Nevertheless, the use of continuous-wave (CW) periodic modulation brings the need for multiple frequency measurements to solve depth ambiguities or cope with multi-path interference. Sequential acquisition of multiple frames reduces the frame rate, while harmonic distortion of the modulation waveforms produces wiggling in the depth estimation. This paper presents and verifies the operation of an IToF CMOS image sensor designed to provide single-shot multi-frequency measurements. A macro-pixel structure allows acquiring multi-frame data in one shot, while resonant demodulation annihilates the harmonic content. The novel architecture consists of $10\,\mu\mathrm{m} \times 10\,\mu\mathrm{m}$ 2-tap pixels with a 20% fill factor (FF). Post-layout simulations show promising 3D reconstruction for up to 16 different simultaneous frequencies.**

***Index Terms*—CMOS image sensor, Time-of-Flight, macro-pixel, single-shot, multi-path.**

I. Introduction

Three-dimensional (3D) sensors are widely used in many areas like mobile phones, machine vision, automotive, and gesture recognition. They can also bring impressive improvements in fields such as safety, industrial automation, virtual and augmented reality, robotics, etc.

3D ranging technologies can be generally divided into optical and non-optical. Radar, geometry-based, and time-of-flight (ToF) scanning systems are among the most commonly used techniques. Among the non-optical sensing methods, the well-known radar range finding system has the advantage of high distance range measurement under various conditions. However, it suffers from low spatial resolution since it requires extensive array size for the receiver [1]. Interferometry, triangulation, and ToF are the three main types of light-based distance sensing methods. These methods rely on the light of wavelengths ranging from 400 to 1000 nanometers (visible and NIR spectrum) [2]. Among these three categories, ToF, despite its lower accuracy, has higher speed and efficiency and is the best candidate for integrated 3D imager implementations. The single-photon avalanche diode (SPAD) is the best option for the direct ToF (DToF) technique in which the sensor directly measures the photon-incidence time by time-to-digital conversion (TDC) [3]. Although this method has many exciting features like long-distance measurement and excellent distance accuracy, it needs special implementation techniques and complex post-processing. Furthermore, TDCs occupy a large area, reducing its spatial resolution [4].

On the other hand, IToF measures the phase difference between the emitted and the received signal, providing high spatial resolution and very simple processing compared to the other counterparts [5], [6]. IToF works based on modulated continuous or pulsed light in concurrence with electrooptical demodulation of a photo-mixing device (lock-in pixels) [7]. Unlike 2D imaging, in IToF, the integration process is controlled by applying predefined signals to the modulation gates. Therefore, depth calculations are based on samples of the cross-correlation of the received signal and the applied control signal. In this way, the range data of each pixel can be retrieved from a group of measurements [8], [9].

Notwithstanding the many advantages IToF offers, it has some critical challenges like motion artifacts, background light error, and ambiguous range. One of the other significant issues in IToF systems is Multi-Path-Interference (MPI), meaning that each pixel may receive more than a single bounce, resulting in a superposition of sinusoidal signals reaching the pixel. MPI errors usually increase the measured distance value since the direct path (desired bounce) distance is shorter than the indirect path (undesired bounce). Typically, the solution for this problem is to increase the number of measurements with different modulating frequencies sequentially, but the resulting increase in the total exposure time decreases the frame rate of depth images. Besides, the required number of frequencies will increase linearly with the number of received bounces. Using the measurements obtained at different frequencies, the distance values for individual bounces can be obtained in a closed-form [10]. In this paper, we demonstrate the successful realization of a single-shot multi-frequency IToF architecture in 180 nm CMOS technology by means of a resonant demodulation concept.

II. Multi-frequency ToF measuring concept

The use of macro-pixels allows for single-shot multi-frequency demodulation [11], thus enabling multiple-path depth imaging at high frame rates. A macro-pixel can be

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 860370. This work was also funded by Spanish Ministry of Science, Innovation and Universities under grant RTI2018-097088-B-C32, from Xunta de Galicia-Consellería de Cultura, Educación e Ordenación Universitaria Accreditation 2019-2022 ED431G-2019/04 and Reference Competitive Group Accreditation 2021-2024 ED431C2021/048, co-funded by (ERDF/FEDER programme).

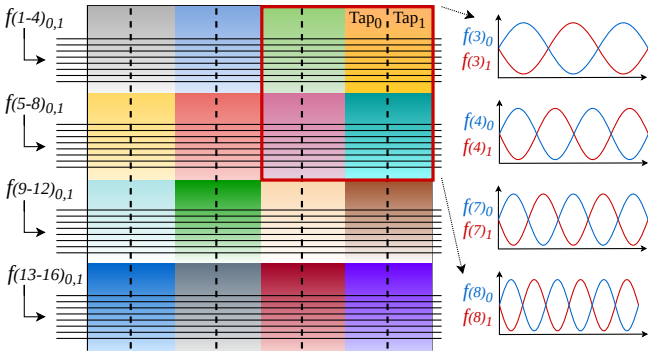

Fig. 1: An array of sub-pixels with 16 possible frequencies forming a single macro-pixel.

formed by grouping numbers of pixels modulated with different frequencies, so-called *subpixels*. In this work, we focus on the case of CW sinusoidal modulation, in which a sinusoidal wave of IR light illuminates the scene at the defined frequency, and the reflected light received by the ToF pixels will generate charges that are accumulated according to the demodulation control signals of the taps. The prevalent 2-tap structure is used in our pixel's circuit.

For illustration, Fig. 1 shows a single macro-pixel consisting of an array of 4 by 4 subpixels. Each subpixel has a 2-tap structure, where both taps demodulate with the same frequency with a 180° phase shift. We denote as $f(i)_{0,1}$ the frequency of the demodulating control signal at subpixel i, with $i = 1, ..., 16$ in this case, for both Tap0 and Tap1. For the sake of clarity, the control signals of the 4 subpixels marked with a red square in Fig. 1 are shown on the right. As can be seen, differently from conventional IToF, the demodulating frequencies are different for each pixel. Each of the 16 subpixels can be either connected to the same frequency (conventional IToF) or to 16 separate frequencies. The entire pixel array, including several macro-pixels, can be formed by repeating this pattern.

In previous work, we presented a resonant demodulation circuit that uses an off-chip inductance connected across the taps to achieve the minimum harmonic distortion [12]. The equivalent circuit for each tap of a subpixel from the modulating gate is modeled as a small capacitor, C_{gate}, connected to resistors, while the whole pixel array can be modeled approximately as two different networks of these capacitors connected in parallel. All pixels of the same type (demodulating at the same frequency) are controlled together, and the parallel-connected control gates are plugged to an inductance of proper value to generate a resonant circuit at the desired frequency. This is summarized in Fig. 2, for an array of n macro-pixels where C_0 and C_1 represent the lumped equivalent capacitances at the gates of the transfer gate transistors TX_0 and TX_1 of the n pixels demodulating at the same frequency, that is, $C_0 = C_1 = n\,C_{\text{gate}}$. For example, for a 4×4 macro-pixel with a different demodulating frequency for all subpixels ($n_{\text{sub}} = 16$), a total of 16 inductances will be required to generate all the frequencies.

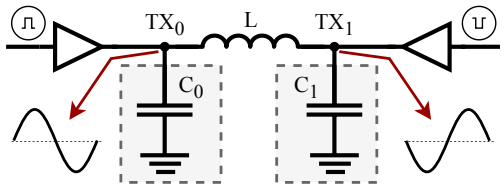

Fig. 2: Simplified electrical model of the resonant demodulation concept. C_0 and C_1 are equivalent capacitance of the whole network connected to each tap.

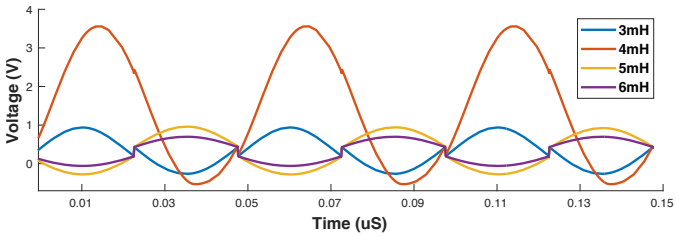

Fig. 3: Resonant demodulation for 4 sub-pixels, connected to a frequency of f = 20 MHZ with different inductance values.

To achieve close-to-zero harmonic distortion, pulse signals with opposite directions were applied to the transfer gates of subpixels through tri-state buffers so that the harmonic distortion reaches the minimum value [13]. For the proper resonant demodulation in the L-C circuit, the duty cycle of the source which drives the tri-state buffer should be sufficiently high. The shape of the resonant demodulation waveform at the gates of the transfer gate transistors TX_0 and TX_1 (Fig. 2) depends on the resistance of the RC network in the whole pixel array. The total average modulation drive power is substantially smaller in this technique than in the standard method.

Fig. 3 shows the parametric simulation results of the electrical modulation waveforms with different inductance values in the frequency of $f = 20\,\text{MHz}$. The simulation is performed for 4 macro-pixels while connecting one subpixel from each macro-pixel (modulating with the same frequency), making a total of 4 connected nodes. The lumped capacitance of the transfer gates of 4 connected subpixels is $C_0 = C_1 = 28.1\,\text{fF}$ (Fig. 2), roughly 4 times the gate capacitance of the individual subpixels ($C_{\text{gate}} = 7\,\text{fF}$), as expected. Next, the corresponding inductance value of around 4 mH for the mentioned frequency and capacitance was analytically calculated for the resonant frequency. A well-defined sine-wave signal is acquired from the LC circuit when the inductance value is near the estimated one for the resonance frequency, as can be seen in Fig. 3. The sinusoidal waveforms from other values of L have lower amplitude and distorted behavior, as expected.

The capability of this system for generating resonant frequencies for modulating the subpixels will be studied experimentally in our future work. However, the simulation results for proving this concept were favorable. In this work, we will focus on the performance of the ToF system with different

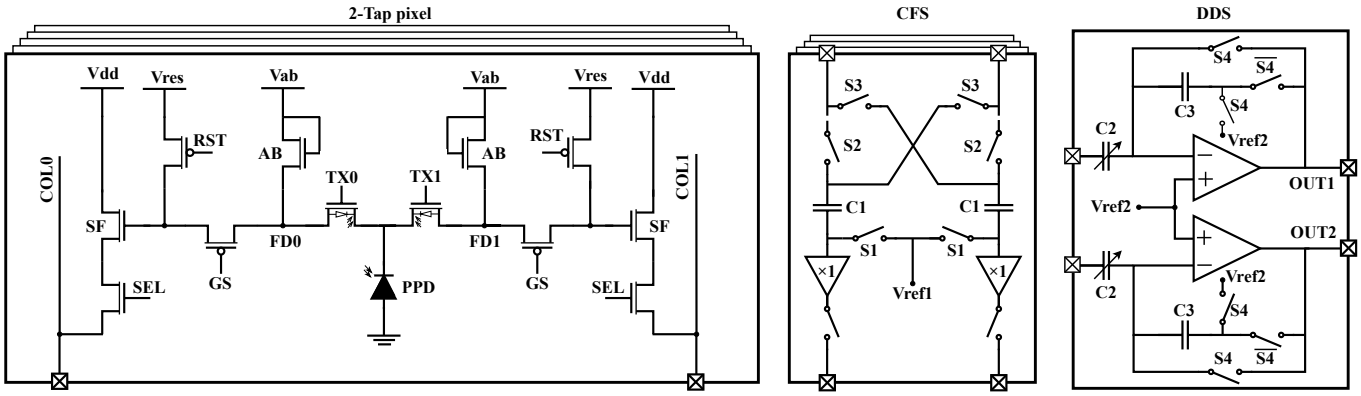

Fig. 4: Schematic of the designed pixel and the readout circuitry.

frequencies in nominal and post-layout simulations.

III. Circuit description

Fig. 4 depicts the circuit schematic of the pixel and the following readout circuitry. In today's implementations, the fabrication of photosensitive parts is usually compatible with CMOS processes. One of the indispensable factors of 3D-sensing pixels is the transfer speed of collected electrons from the photodetector to the transfer gate. Photogates are more common photodetectors due to their intrinsically higher demodulation speed and charge transfer. Pinned photodiodes (PPD), on the other hand, have higher quantum efficiency and depth resolution. Small PPDs can benefit from the fringe field at their edge. However, the larger a pixel size gets the less benefit from the fringe field. Therefore, in PPDs, speed issues are more likely for a large pixel. The process techniques can be modified by creating an internal electrical field to increase the charge transfer speed to solve this problem. As another option, layout modifications that reduce the travel distance for electrons can be used. Also, both approaches can be combined simultaneously by generating an internal electrical field and decreasing distance. In this work, the layout of the PPD has relatively small dimensions ($4\,\mu\mathrm{m} \times 4\,\mu\mathrm{m}$) and an extended poly-silicon layer on top to increase the charge demodulation speed.

The design includes a symmetrical 2-tap structure with a PPD as a photosensitive part and reset, global shutter, source follower (SF) and row select transistors as a buffer and reading part [14]. The architecture consists of array of $10\,\mu\mathrm{m} \times 10\,\mu\mathrm{m}$ pixels with a 20% fill factor (FF). The PMOS transistors are selected to perform a hard reset to optimize signal swing and avoid image delays. For the SF and row select, low threshold NMOS transistors are selected to enable a larger output voltage swing. Also, a relatively big gate area is selected for the SF to minimize the noise, even though it reduces the fill factor. The analog supply voltage (Vdd) is 3.3 V for the pixel for a higher signal dynamic range.

The following stages after the exposure time are the Column Filter Stage (CFS) and Double-Delta-Sampling (DDS) blocks. The CFS block can process and filter pixel signals, and it consists of two decoupling capacitors followed by voltage followers whose outputs can be selected by a column decoder to be sent sequentially through the DDS output stage.

The DDS schematic is shown in Fig. 4-right. It amplifies the pixel output and sets the desired offset voltage. The final output of the system will be $V_{\mathrm{ref}_2} \pm \left(\frac{C2}{C3}\right) \Delta V$, where ΔV is the voltage difference between the two outputs of the circuit. The two-stage opamp is designed based on the class-AB output stage used in [15] and has a high gain and output swing. The first stage is a folded cascode section to increase the value of the DC voltage gain of the opamp. A class-AB output stage with a common-source configuration is used as the second part, providing a rail-to-rail signal swing. The width and length of the transistors are adequately optimized to reach 130 dB gain, 32 MHz gain-bandwidth (GBW), and 33 V/μs of slew-rate (SR) in transient with 15 pF load capacitance, while the design provides desirable stability.

IV. Simulation Results and Discussion

The active pixel sensor, presented in Section III operates in reset, integration, and readout stages to retrieve the 3D data. In the 2-tap structure, at least two cycles of measurements are required to calculate the three unknown parameters of the received signal (amplitude, offset, phase shift).

The phase shift of the received signal can be calculated using the differential output from the two frames using (1):

$$\Delta\phi = \arctan\left(\frac{V_{270^\circ} - V_{90^\circ}}{V_{180^\circ} - V_{0^\circ}}\right) \tag{1}$$

where $V_{0^\circ}, ..., V_{270^\circ}$ are the outputs of the subpixel for 0°, 90°, 180°, and 270° phase shifts of the demodulating gates' control signal with respect to the emitted light. With this technique, the phase shift is calculated using different phase shifts (0°/180° for the first frame and 90°/270° for the second frame) between the illumination control signal and the pixel modulation signals. To test the performance of the pixel, we carried out electrical simulations in a CIS 180 nm CMOS technology, operating at 3.3 V and 20 MHz. First, we measured the differential outputs of the first ($V_{180^\circ} - V_{0^\circ}$) and second frames ($V_{270^\circ} - V_{90^\circ}$) for different relative phase differences

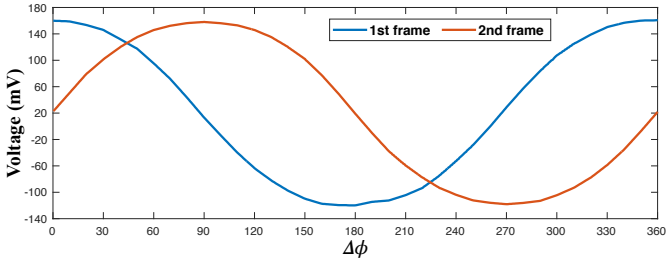

Fig. 5: Electrical simulation showing the nominal output voltage difference as a function of the phase shift at 20MHz for the first and second frames.

between the emitted and the received light signals. The simulated circuit includes the pixel, the CFS, the DDS, and the parametric simulations performed without any amplification in the DDS stage. Fig. 5 shows the simulation results. The system's outputs are samples of a cross-correlation function between the incoming light signal and the demodulation signals controlling the gates. Because both of these signals are sinusoidal, the resulting cross-correlation is also sinusoidal, as can be seen. The outputs from the frames have different values with respect to frequency and can result in distinct proportion of $\left(\frac{V_{270^\circ}-V_{90^\circ}}{V_{180^\circ}-V_{0^\circ}}\right)$. The distance of a single bounce then can be retrieved from:

$$d = \frac{c\Delta\phi}{4\pi f}\ , \tag{2}$$

where c is the speed of light, $\Delta\phi$ is the phase shift, and f is the resonant frequency.

In our second test, we performed a post-layout simulation for an 8 by 8 array of subpixels modulating at different frequencies from 8 MHz to 40 MHz (2 MHz increment), forming 4 macro-pixels of 4 by 4 subpixels each. The purpose of this post-layout simulation is to demonstrate the single-shot multi-frequency demodulation concept by applying a different modulating frequency to each of the 16 pixels forming the macro-pixel. The same 16 frequencies are used in the 4 macro-pixels. For this simulation, a scene with the capital F letter with 4 vertices and 4 pixels in the right and left borders is considered as an example. The ground truth distance considered for the F letter, the vertices, the borders, and the background is 0.5, 1, 1.5, and 2.5 meters, respectively. First, we calculated the corresponding phase shift for each subpixel concerning its frequency and shifted the input light signal accordingly. As it has been explained in Section II, for each macro-pixel, we apply 16 different frequencies. Then, we calculated the distance according to (2). Fig. 6 shows the post-layout simulation results after retrieving the 3D information. The results match the ground truth depth (Fig. 6-top right) well, showing the capability of our system to work with different demodulation frequencies.

Fig. 7 shows the layout design of the entire ToF range imaging system, which consists of an array of pixels, the modulation control signal tree, the biasing circuitry, the row

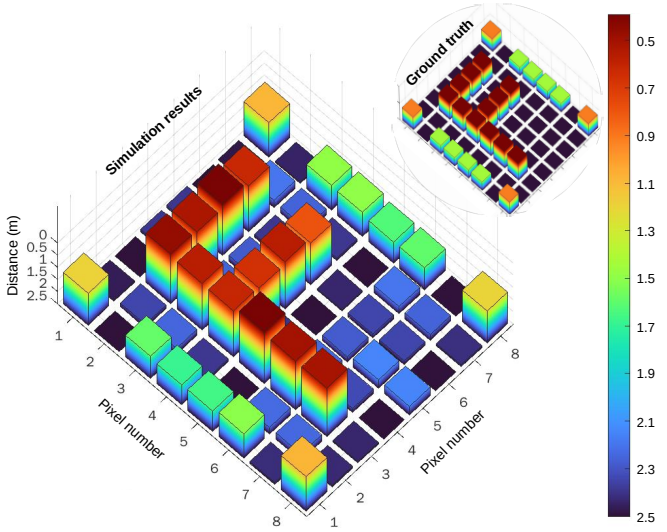

Fig. 6: Retrieved distance of an example from post-layout simulation results of the 8 by 8 array of pixels and readout circuitry vs. ground truth (top-right inset).

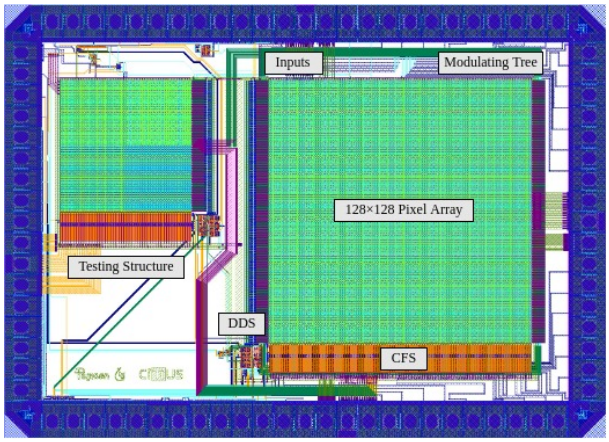

Fig. 7: System-Level architecture and layout of the single-shot multi-frequency IToF system with its parts labeled.

and column decoders, the CFS, and the DDS stage. The system has an array-shared modulation signal generator, row and column control circuits, driver circuits, and bias circuits for in-pixel operational amplifiers.

V. Conclusion

This paper presents the post-layout simulation results of an 8 by 8 array of IToF pixels with readout circuitry. The proposed macro-pixel-based architecture allows for realizing the idea of single-shot multi-frequency demodulation. Furthermore, perfect sinusoidal control signals at the demodulation gates, arising from a resonant construction, enable native harmonic cancellation per pixel or group of pixels. Resonant demodulation can further reduce the average drive current of pixels, resulting in considerable power-saving and paving the way for low-powered 3D imaging systems.

REFERENCES

[1] T.-H. Hsu and et al, "A CMOS time-of-flight depth image sensor with in-pixel background light cancellation and phase shifting readout technique," *IEEE Journal of Solid-State Circuits*, vol. 53, no. 10, 2018.

[2] R. Lange and P. Seitz, "Solid-state time-of-flight range camera," *IEEE Journal of Quantum Electronics*, vol. 37, no. 3, 2001.

[3] K. Morimoto and et al., "Megapixel time-gated SPAD image sensor for 2D and 3D imaging applications." *Optica*, vol. 7, no. 4, 2020.

[4] K. Seong-Jin and et al, "A CMOS image sensor based on unified pixel architecture with time-division multiplexing scheme for color and depth image acquisition," *IEEE Journal of the Electron Devices Society*, vol. 47, no. 11, pp. 2834–2845, 2012.

[5] Y. Shirakawa, K. Yasutomi, K. Kagawa, S. Aoyama, and S. Kawahito, "An 8-Tap CMOS Lock-In Pixel Image Sensor for Short-Pulse Time-of-Flight Measurements," *Sensors*, vol. 20, no. 4, 2020.

[6] K. Kagawa, T. Kokado, Y. Sato, F. Mochizuki, H. Nagahara, T. Takasawa, K. Yasutomi, and S. Kawahito, "Multi-tap macro-pixel based compressive ultra-high-speed cmos image sensor," in *Proceedings of 2019 the International Image Sensor Workshop (IISW)*, 2019, pp. 270–273.

[7] A. Bhandari, M. Heredia Conde, and O. Loffeld, "One-Bit Time-Resolved Imaging," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 42, pp. 1630–1641, 2020.

[8] M. Heredia Conde, *Compressive sensing for the photonic mixer device.* Springer, 2017.

[9] T. Möller and et al., "Robust 3D measurement with PMD sensors." *Range Imaging Day, Zürich*, vol. 7, no. 8, 2005.

[10] M. Heredia Conde, T. Kerstein, B. Buxbaum, and O. Loffeld, "Fast multipath estimation for PMD sensors," in *5th International Workshop on Compressed Sensing Theory and its Applications to Radar, Sonar, and Remote Sensing (CoSeRa 2018)*, 2018.

[11] M. Heredia Conde, K. Kagawa, T. Kokado, S. Kawahito, and O. Loffeld, "Single-Shot Real-Time Multiple-Path Time-of-Flight Depth Imaging for Multi-Aperture and Macro-Pixel Sensors," in *2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2020, pp. 1469–1473.

[12] P. F. Shahandashti, P. Lopez, V. Brea, D. Garcia-Lesta, and M. Heredia Conde, "Proposal of a Single-Shot Multi-Frame Multi-Frequency CMOS ToF Sensor," in *2021 28th IEEE International Conference on Electronics, Circuits and Systems (ICECS)*, 2021.

[13] R. M. Conroy, A. A. Dorrington, and A. D. Payne, "A Power Saving Modulation Technique for Time-of-Flight Range Imaging Sensors," in *Proceedings of the SPIE. Videometrics, Range Imaging, and Applications*, vol. 8085, 2011.

[14] D. Stoppa and et al., "A Range Image Sensor Based on 10-μm Lock-In Pixels in 0.18-μ m CMOS Imaging Technology," *IEEE Journal of Solid-State Circuits*, vol. 46, no. 1, 2010.

[15] O. Pereira-Rial and et al, "Compact CMOS Class-AB Output Stage With Robust Behavior Against PVT Variations," in *Proceedings of the 27th IEEE International Conference on Electronics, Circuits and Systems (ICECS)*, 2020.