

# A Half-Pulse 2-Tap Indirect Time-of-Flight Ranging Method with Sub-Frame Operation for Depth Precision Enhancement and Motion Artifact Suppression

Chia-Chi Kuo<sup>1</sup>, Rihito Kuroda<sup>1,2</sup>

<sup>1</sup> Graduate School of Engineering, Tohoku University,  
6-6-05, Aza-Aoba, Aramaki, Aoba-ku, Sendai, Miyagi, Japan 980-8579

<sup>2</sup> New Industry Creation Hatchery Center, Tohoku University

TEL: +81-22-795-4833, Email addresses: kuo.chiachi.s2@dc.tohoku.ac.jp, rihito.kuroda.e3@tohoku.ac.jp

## ABSTRACT

This paper presents a new 2-tap indirect time-of-flight (iToF) ranging method using half-pulse (HP) modulation and sub-frame 4-phase sampling. The proposed operation and derived theoretical depth noise equations were verified by the experiments, which achieve >25% and >29% depth precision enhancement for 0.4-3 m range with proposed HP1 and HP2 method, respectively. The motion artifact suppression is also demonstrated using sub-frame operation that provides clear depth images for a scene with a moving object.

## INTRODUCTION

3-D imaging system has been developed rapidly and applied to various fields. In these days, there are more emerging opportunities that can make use of this technology such as autonomous vehicles, industrial automation, computer-human interaction, and so on. To provide a reliable spatial information, a high performance 3-D camera is desired. The 2-tap indirect time-of-flight (iToF) based depth image sensors have been a popular choice that can realize a range imager with high resolution, low power consumption, low GPU computation complexity and good system compactness, simultaneously [1], [2].

To achieve a better ranging precision, increasing the system signal-to-noise ratio (SNR) and the modulation frequency have always been the key targets, which were revealed to be inversely proportional to the depth noise [3]. However, the existed trade-off between the imaging performances should be taken into concerns for the realization of an iToF system.

Firstly, a higher full-well capacity (FWC) design can obtain a higher SNR whereas a larger pixel size and longer exposure period are required. Secondly, due to the phase wrapping, the unambiguous range will reduce while applying a higher modulation frequency. Although the multi-frequency synthesis had been reported to be useful to extend the valid scope range [4], ensuring the robustness of phase unwrapping and a good demodulation contrast (DC) under high-speed modulation remain a challenge [4], [5]. Besides, the intensive power consumption during the exposure period, which mainly lead by the toggling demodulation gates and the emitted light, may reduce the robustness and durability of a ranging system.

On the other hand, the conventional 2-tap 4-phase (2T-4PH) operation requires two successive frames to reconstruct a background light (BGL) canceled depth

image [6]. However, the requirement of temporal multiplexing will induce an increase of motion artifact due to the frame-to-frame processing latency.

In this article, a new 2-tap ranging method using half-pulse (HP) modulation with sub-frame 4-phase sampling is introduced to address these issues.

## OPERATION PRINCIPLE

Fig.1 shows the ranging methodologies for 2T-4PH iToF system using the conventional continuous square-pulse (SP) [7] and proposed HP modulation, where  $c$  is the speed of light. The modulation frequency, cycle time, and pulse width of SP and HP are denoted by  $f_m$ ,  $T_C$ ,  $T_{SP}$  and  $T_{HP}$ , respectively.

The received light that demodulated by Tap1 and Tap2 during two exposure periods,  $PH(0, \pi/2)$ , are denoted by  $Q1(0, \pi/2)$  and  $Q2(0, \pi/2)$ , respectively. Note that the emitted light amplitude of HP is doubled from SP to have an equal averaged light intensity during the modulation.

For the distance ( $d$ ) calculation, Fig. 2 shows the equations to obtain the light traveling time ( $T_{TOF}$ ), where  $R$  is the time-shift ratio which is defined from 0 to 4 through the time-window, TW(1) to TW(4).

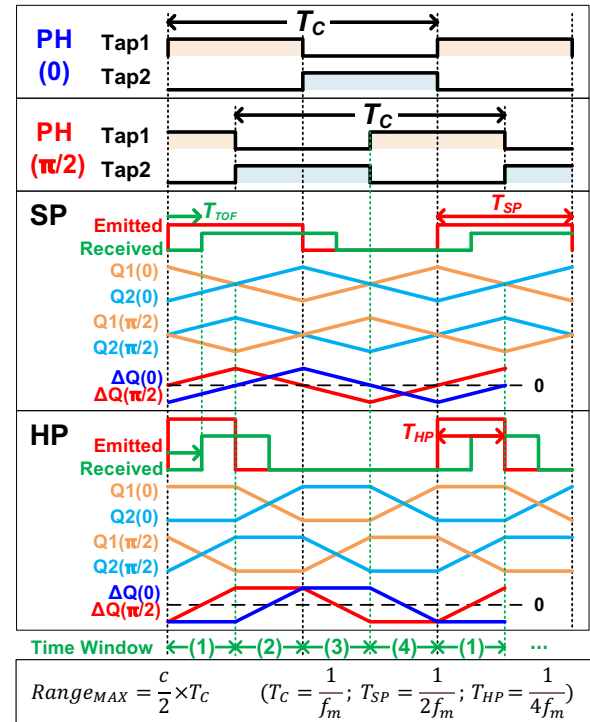


Fig. 1 Operation principle of 2T-4PH ranging methods using SP and HP modulation.

$$\text{Dist.}(d) = \frac{c}{2} \times T_{ToF}; \quad T_{ToF} = T_C \times \frac{R}{4} = \frac{R}{4f_m} \quad (0 \leq R < 4)$$

**Square-Pulse Modulation - SP**

$$R_{SP} = \begin{cases} 1 + \frac{\Delta Q_0}{|\Delta Q_0| + |\Delta Q_{\pi/2}|}, & \text{if } \Delta Q_{\pi/2} \geq 0 \\ 3 + \frac{-\Delta Q_0}{|\Delta Q_0| + |\Delta Q_{\pi/2}|}, & \text{if } \Delta Q_{\pi/2} < 0 \end{cases}$$

**Half-Pulse Modulation - HP1**

$$R_{HP1} = \begin{cases} 0 + \frac{Q_{2\pi/2} - Q_{2Q_0}}{Q_{1\pi/2} + Q_{2\pi/2} - 2 \cdot Q_{2Q_0}} & @TW(1) \\ 1 + \frac{Q_{2Q_0} - Q_{1\pi/2}}{Q_{1_0} + Q_{2Q_0} - 2 \cdot Q_{1\pi/2}} & @TW(2) \\ 2 + \frac{Q_{1\pi/2} - Q_{1_0}}{Q_{1\pi/2} + Q_{2\pi/2} - 2 \cdot Q_{1_0}} & @TW(3) \\ 3 + \frac{Q_{1_0} - Q_{2\pi/2}}{Q_{1_0} + Q_{2Q_0} - 2 \cdot Q_{2\pi/2}} & @TW(4) \end{cases}$$

**Half-Pulse Modulation - HP2**

$$R_{HP2} = \begin{cases} \frac{1}{2} \cdot \left(1 - \frac{\Delta Q_{\pi/2}}{\Delta Q_0}\right) & @TW(1) \\ \frac{1}{2} \cdot \left(3 + \frac{\Delta Q_0}{\Delta Q_{\pi/2}}\right) & @TW(2) \\ \frac{1}{2} \cdot \left(5 - \frac{\Delta Q_{\pi/2}}{\Delta Q_0}\right) & @TW(3) \\ \frac{1}{2} \cdot \left(7 + \frac{\Delta Q_0}{\Delta Q_{\pi/2}}\right) & @TW(4) \end{cases}$$

Fig. 2 Time-shift ratio equations for depth calculation.

$$TW = \begin{cases} (1), & \text{if } SQ < 0, DQ < 0 \\ (2), & \text{if } SQ > 0, DQ < 0 \\ (3), & \text{if } SQ > 0, DQ > 0 \\ (4), & \text{if } SQ < 0, DQ > 0 \end{cases}$$

Fig. 3 Time-window logic for HP ranging methods.

The time-shift ratio ( $R$ ) obtained by SP, HP1 and HP2 ranging methods are denoted by  $R_{SP}$ ,  $R_{HP2}$  and  $R_{HP1}$ , respectively. Note that SP and HP2 use the differential demodulated signals from  $PH(0, \pi/2)$ , which are declared as  $\Delta Q_0 (= Q_{2Q_0} - Q_{1_0})$  and  $\Delta Q_{\pi/2} (= Q_{2\pi/2} - Q_{1\pi/2})$ , whereas the HP1 uses three individual signals to perform the BGL cancelling (BGLC) scheme for depth calculation.

For the HP modulation, the logic shown in Fig. 3 is used to determine the TW of the ranging result, where the summation of phase  $SQ (= \Delta Q_0 + \Delta Q_{\pi/2})$  and difference of phase  $DQ (= \Delta Q_0 - \Delta Q_{\pi/2})$  are created.

By applying the propagation of errors [8], the theoretical depth noise ( $\sigma_d$ ) equations were derived and expressed in Fig. 4, where DC is the demodulation contrast and  $R_S = N_{ToF}/N_S$ . The number of total electrons integrated in a unit pixel during a single frame from the emitted light and BGL are denoted by  $N_S$  and  $N_{BT}$ .  $N_{ToF}$  is the number of demodulated electrons from  $N_S$  in the single tap, which will increase along with  $T_{ToF}$  in each time-window, where  $0 \leq R_S < 1$ . The readout noise referred to pixel floating diffusion (FD) is denoted by RN.

$$\text{Depth Noise } (\sigma_d) = \frac{c}{8f_m} \times \frac{\sigma_R}{DC}$$

$$\sigma_{R_{SP}} = \frac{\sqrt{(N_S + N_{BT} + 2RN^2)(1 - 4R_S + 8R_S^2)}}{N_S}$$

$$\sigma_{R_{HP1}} = \frac{\sqrt{N_\phi(1 - R_S) + (N_{BT} + 2RN^2)(1 - 3R_S + 3R_S^2)}}{N_S}$$

$$\sigma_{R_{HP2}} = \frac{\sqrt{\left(\frac{N_S}{2} + \frac{N_{BT}}{2} + RN^2\right)(1 - 2R_S + 2R_S^2)}}{N_S}$$

Fig. 4 Theoretical depth noise equations.

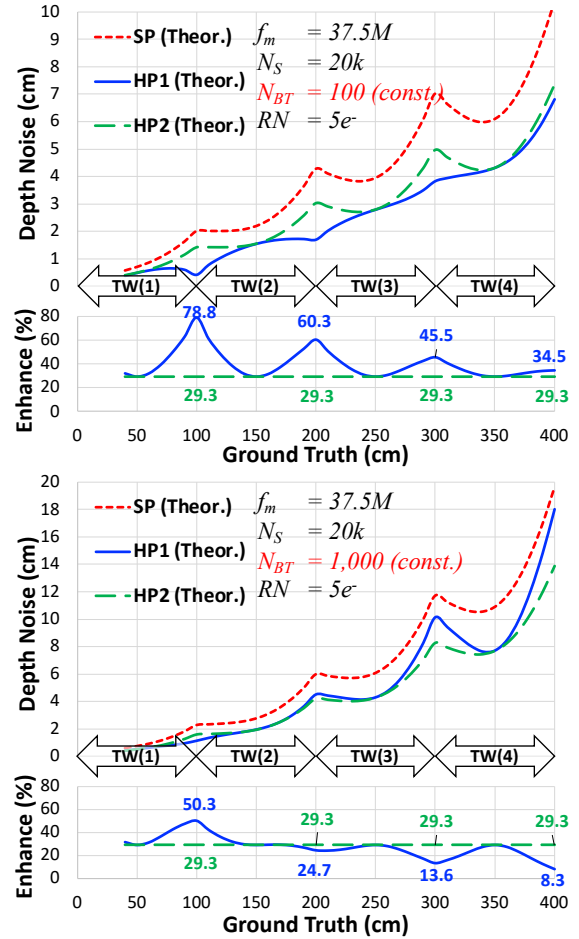


Fig. 5 Theoretical depth noise of the ranging methods.

Fig. 5 compares the theoretical depth noise curves of SP, HP1 and HP2 ranging method. Note that  $N_S$  decreased rapidly as a function of distance squared, whereas  $N_{BT}$  was a constant for all range. The depth noise enhancement was obtained by the ratio of HP to SP and depicted in percentage.

With a weak influence from BGL, HP1 method can improve the depth precision efficiently, especially when  $R_S$  is close to 0 or 1 at each TW. On the contrary, HP2 method guarantees a consistent noise reduction capability for all conditions. In addition, the reported  $\Delta$ -INT BGLC scheme can also be implemented using HP2 against a stronger ambient light environment [9].

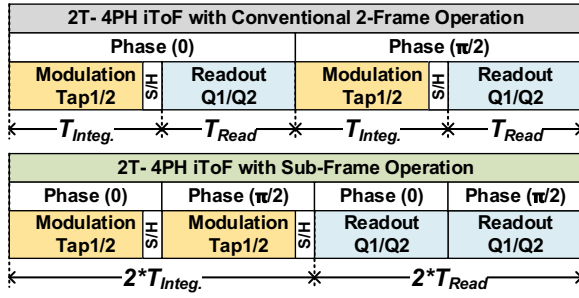


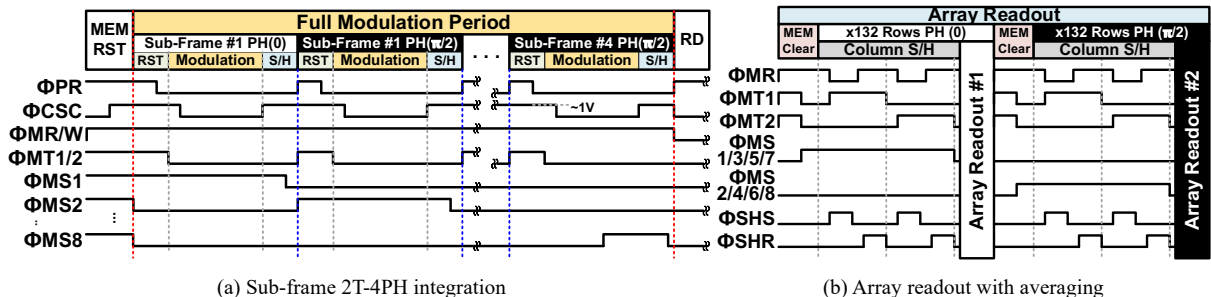
Fig. 5 Concept of sub-frame 2T-4PH operation.

To perform the BGLC algorithms on a 2T-4PH iToF sensor, two modulation periods followed by the array readout are required. As drawn in Fig. 5, the conventional 2-frame operation will encounter with a considerable information loss between  $PH(0, \pi/2)$ , particularly for a high pixel count sensor. In contrast, by applying the sub-frame operation, a reduction of motion artifact can be expected owing to the compact modulation period. Subsequently, the array readout will be carried out after the modulations.

### PIXEL & TIMING DIAGRAM

A reported 4-tap iToF imager [10], shown in Fig. 6, was used to evaluate the proposed techniques. By applying the adjusted timing drawn in Fig. 7, this sensor can be operated as a 2T-4PH iToF imager. The equivalent pixel circuit diagram is shown in Fig. 8, which consists of a high-speed charge collection photodiode (PD), demodulation gates (TGs), buried channel source follower (PSF), current source (PCS), cascode switch (CSC), auto-zeroing capacitor ( $C_{AZ}$ ) and  $2 \times 8$  1-T 1-C analog memory ( $C_{MEM}$ ) with control devices (MW, MRST, MTs, MSs), which share a column readout buffer (MSF).

Fig. 7 depicts the detailed timing diagram. The full integration period is constructed by  $2 \times 4$  subframes, where  $PH(0, \pi/2)$  modulations are performed and sampled into the memory array alternatively. The auto-zeroing sampling is adopted to eliminate the reset thermal noise. Before the column sampling, the charge-domain binning is applied to each phase by mixing the signal charges in the corresponded 4-phase memories. Owing to the averaging effect, the photon shot noise, noise from pixel transistors and kTC noise of the memory can be reduced. Finally, the signals in  $PH(0, \pi/2)$  are readout in sequence.



(a) Sub-frame 2T-4PH integration

(b) Array readout with averaging

Fig. 7 Circuit timing diagram of sub-frame 2T-4PH operation.

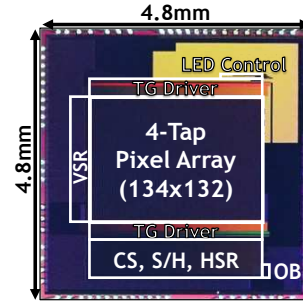


Fig. 6 Chip micrograph [10].

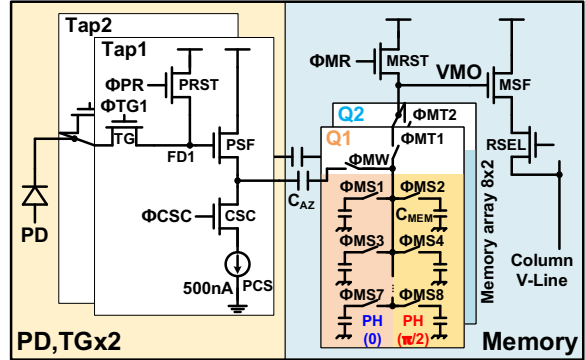


Fig. 8 The equivalent 2T-4PH iToF pixel structure with in-pixel memory array.

### EXPERIMENTAL RESULTS

The basic sensor characteristics were measured beforehand. A single-tap FWC around  $12ke^-$ , and a readout noise floor of  $5.4e^-$  with 4-subframe averaging were confirmed. The DC of 85% with 20 ns demodulation pulse width was obtained, where the modulation light was generated by an 850 nm VCSEL. The following depth performances were characterized by the center  $10 \times 10$  pixels over 100 consecutive frames with a F/1.4 lens and IR bandpass filter.

This sensor was operated at a modulation frequency of 25 MHz, which corresponds to an unambiguous range of 6 m for 4-TW referred to Fig. 1. However, due to the circuit timing constraint, the measurable range was confined to 1.5 m with TW(1) using the conventional continuous SP modulation, and to 3.0 m with TW(1)&(2) using the proposed HP modulation.

As the experimental data shown in Fig. 9, the depth precision was enhanced by 44% and 31% at 1.5 m using the HP1 and HP2 ranging method, respectively.

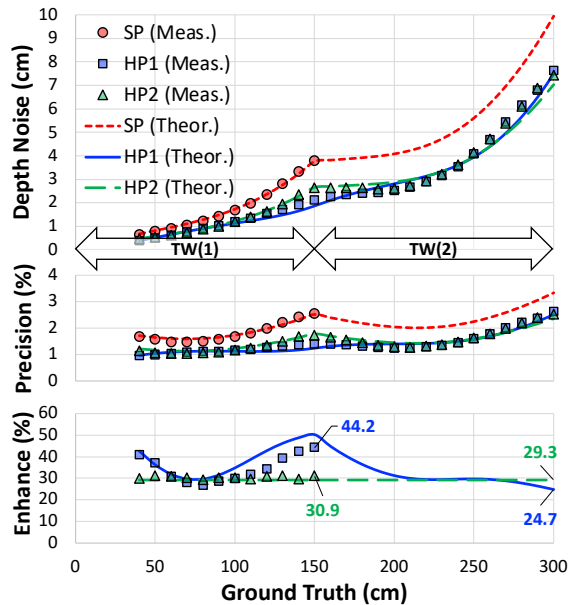


Fig. 9 Experimental and theoretical depth noise curves.

The experimental results show a good agreement with the theoretical calculations in Fig. 4, which can be adopted to estimate the depth noise for the 2-tap iToF ranging methods. As can be seen, HP1 and HP2 methods can provide  $>25\%$  and  $>29\%$  precision enhancement, respectively, for the range of 0.4-3 m.

Besides, by giving a proper sensor design with a  $f_m$  of 37.5 MHz,  $N_s$  of 20,000, DC of 90% and RN of  $< 5e^{-}$ , which are achievable values in a state of art iToF system design [11], a high precision of  $<1\%$  error for a range of 0.4-4m is to be available using HP1 method.

Fig. 10 demonstrated the sample images captured under 30 fps. The image distortion on the hand movement was observed while using the conventional 2-frame operation with SP modulation. In contrast, higher quality depth images with lower noise and suppressed motion artifact were reconstructed using the proposed HP methods with sub-frame operation.

## CONCLUSION

In this paper, a 2-tap 4-phase iToF ranging method using HP modulation with sub-frame operation were presented. Both HP1 and HP2 methods can provide a better ranging precision and be utilized in different ways. For the indoor applications, HP1 method is recommended that can achieve the highest depth precision. On the other hand, HP2 has an advantage in a lower SNR ranging system, such as strong ambient light environment and long-distance ranging.

The proposed HP methods can be adopted in exhibited 2-tap iToF sensors to improve their ranging performance without increasing the modulation period or frequency. This indicates a higher framerate and lower power consumption can be expected. Also, to ensure the depth imaging quality with moving targets, the motion artifact suppression was demonstrated by applying the sub-frame 4-phase sampling.

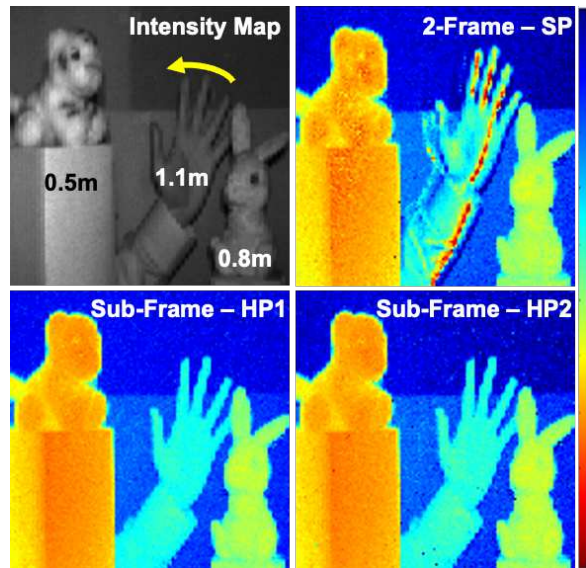


Fig. 10 Captured sample images with hand movement.

The developed technologies show a promising potential to enhance the performance and reliability for various 3-D imaging applications.

## REFERENCES

- [1] C. S. Bamji *et al.*, "iMpixel 65nm BSI 320MHz demodulated TOF Image sensor with  $3\mu\text{m}$  global shutter pixels and analog binning," in *2018 IEEE International Solid-State Circuits Conference (ISSCC)*, 2018, pp. 94–96.
- [2] Y. Ebiko *et al.*, "Low power consumption and high resolution 1280X960 Gate Assisted Photonic Demodulator pixel for indirect Time of flight," in *2020 IEEE International Electron Devices Meeting (IEDM)*, 2020, pp. 31–33.
- [3] Y. Kato *et al.*, "320×240 Back-Illuminated 10- $\mu\text{m}$  CAPD Pixels for High-Speed Modulation Time-of-Flight CMOS Image Sensor," *IEEE J Solid-State Circuits*, vol. 53, no. 4, pp. 1071–1078, Apr. 2018.
- [4] C. S. Bamji *et al.*, "A 0.13  $\mu\text{m}$  CMOS System-on-Chip for a  $512 \times 424$  Time-of-Flight Image Sensor with Multi-Frequency Photo-Demodulation up to 130 MHz and 2 GS/s ADC," *IEEE J Solid-State Circuits*, vol. 50, no. 1, pp. 303–319, Nov. 2015.
- [5] C. Bamji *et al.*, "A Review of Indirect Time-of-Flight Technologies," *IEEE Trans Electron Devices*, vol. 69, no. 6, pp. 2779–2793, Jun. 2022.
- [6] D. Stoppa *et al.*, "A range image sensor based on 10- $\mu\text{m}$  lock-in pixels in 0.18- $\mu\text{m}$  CMOS imaging technology," in *IEEE Journal of Solid-State Circuits*, Jan. 2011, vol. 46, no. 1, pp. 248–258.
- [7] J. Cho *et al.*, "A 3-D camera with adaptable background light suppression using pixel-binning and super-resolution," *IEEE J Solid-State Circuits*, vol. 49, no. 10, pp. 2319–2332, 2014.
- [8] R. Lange and P. Seitz, "Solid-state time-of-flight range camera," *IEEE J Quantum Electron*, vol. 37, no. 3, pp. 390–397, 2001.
- [9] D. Kim *et al.*, "Indirect time-of-flight CMOS image sensor with on-chip background light cancelling and pseudo-four-tap/two-tap hybrid imaging for motion artifact suppression," *IEEE J Solid-State Circuits*, vol. 55, no. 11, pp. 2849–2865, 2020.
- [10] C.-C. Kuo and R. Kuroda, "A 4-Tap CMOS Time-of-Flight Image Sensor with In-pixel Analog Memory Array Achieving 10Kfps High-Speed Range Imaging and Depth Precision Enhancement," in *2022 IEEE Symposium on VLSI Technology and Circuits (VLSI Technology and Circuits)*, 2022, pp. 48–49.
- [11] Y. Kwon *et al.*, "A 2.8  $\mu\text{m}$  pixel for time of flight CMOS image sensor with 20 ke-full-well capacity in a tap and 36% quantum efficiency at 940 nm wavelength," in *2020 IEEE International Electron Devices Meeting (IEDM)*, 2020, pp. 32–33.

# A 3.5 $\mu$ m Indirect Time-of-Flight Pixel with In-Pixel CDS and 4-Frame Voltage Domain Storage

Erez Tadmor<sup>1</sup>, Ben Dror<sup>1</sup>, Guy Likver<sup>1</sup>, Gal Fadida<sup>1</sup>, Zvika Veig<sup>1</sup>, Seiji Takeuchi<sup>2</sup>, Toshiaki Rai<sup>2</sup>, Atsushi Noda<sup>2</sup>

onsemi, <sup>1</sup> Netiv Haor 1, Haifa, Israel, [erez.tadmor@onsemi.com](mailto:erez.tadmor@onsemi.com); <sup>2</sup> Gifu, Japan

**This paper presents a new 3.5 $\mu$ m indirect Time-of-Flight (iToF) pixel with in-pixel CDS circuit, ability to store 4 frames in voltage-domain storage capacitors, 38% QE @ 940nm, >25ke- linear full well, and up to 200MHz modulation frequency with >95% modulation contrast. The ability to store all the required raw data for depth calculation inside the pixel enables depth sensing with low motion artifacts and simplifies in-chip depth calculation.**

## Motivation

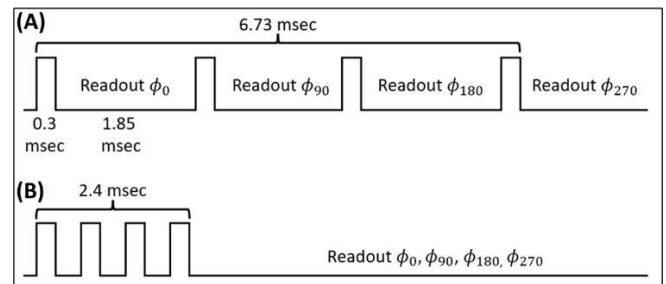
iToF cameras usually require 4-8 exposures with different phase shifts between the illumination and the pixel modulation in order to reconstruct a depth scene. [1,2] The data from these exposures normally has to be read out and stored externally to the sensor. This makes iToF cameras prone to motion artifacts due to coupling between total exposure time and readout time, as shown in Fig.1. In addition, several external frame buffers are required in the system level to store the phase data until all the data is ready for calculation. These problems get worse with increasing resolution, as readout time takes longer, and required frame buffers become larger. In this work we present a pixel that solves this problem using in-pixel voltage domain storage of up to 4 frames.

## Pixel Architecture – In-Pixel CDS & Storage

The pixel consists of a pinned photodiode with two symmetrical modulation gates and a global shutter gate. The modulation gates control the electrostatic potential in the photodiode to quickly sweep electrons into one of the two charge-domain storage gates where they are stored for the duration of the exposure period. The global shutter enables operating in pulsed or hybrid modes, where the modulation is not continuous, but is done in short bursts. When coupled with high peak-power illuminator, this improves the system performance under ambient light [3].

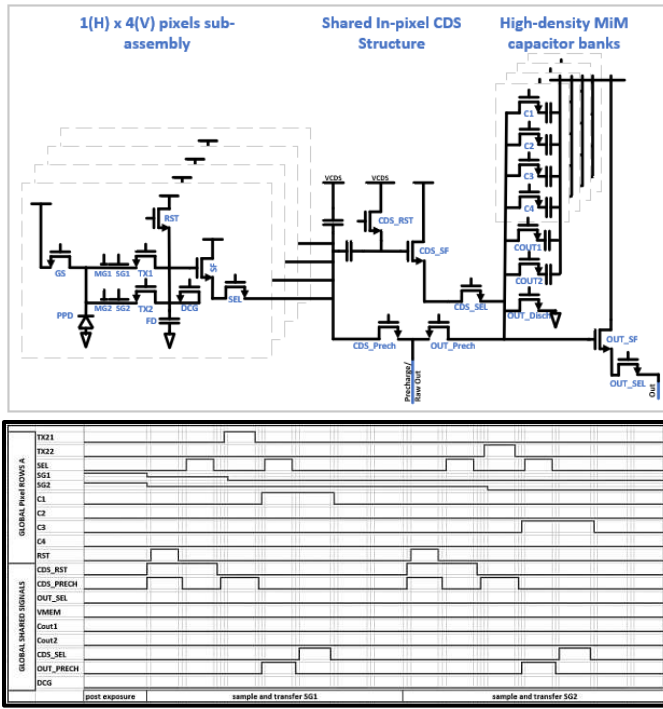
Post-exposure, charge can be read out from the storage gates in a correlated double sampling (CDS) operation. The readout can be performed in the standard way, row wise and out of the sensor, or using the in-pixel CDS and into voltage domain storage capacitors. The mechanism that allows the latter option is presented in Fig.2(A): Each group of 4 pixels share a sample and hold circuit that is then connected to 16 MIM capacitors with high capacitive density. This way, the entire 1.2MP array can be sampled into the storage capacitors in under 400 microseconds.

This operation quickly removes the signal from the charge domain, where it is vulnerable to parasitic light, into the voltage domain. It reduces the effective exposure time by 65%-90% (Depending on exposure time and readout time ratio), which means that the motion artifacts described previously are practically eliminated. Finally, all the data required for depth calculation is being stored in the pixel level so external frame buffers become redundant. It therefore simplifies the depth reconstruction pipeline, as all the data from a specific pixel is transmitted out in the order needed for efficient calculation. This permits depth calculation to be done in the sensor, using very simple and efficient pipelined logic.



**Figure 1(A) – Illustration of the effective exposure time (first-to-last photon time) in a hypothetical 1280x960 iToF sensor, with 4 MIPI lanes outputting 12 bits per pixel at 2Gbps, and exposure time of 300 $\mu$ s per phase, assuming only differential data is read from the sensor. (B) - Using in-pixel CDS and storage, the effective exposure time is reduced by 65% which will result in significantly lower motion artifacts.**

The basic operation of the in-pixel CDS mechanism is described in Figure 2(B). After one cycle of exposure is finished signal is stored in the pixel storage gates (SG). The SG bias is kept at an intermediate level to achieve optimal tradeoff between linear full well and dark current generation. First, the FD is reset, and its value is sampled on the left side of the sample-and-hold capacitor, while right side of the capacitor is being reset. In the next step, the right side of the sample-and-hold capacitor is disconnected from supply by CDS\_RST, and charge from SG1 is transferred to FD, and sampled on the capacitor, now holding the reset subtracted signal. The subtracted signal is then sampled through the second SF to the storage capacitor C1. The operation is repeated for sampling SG2 and the signal is stored on C3. This operation is repeated four times as every four pixels share a CDS circuit.



**Figure 2(A) - A schematic drawing of 4 pixels sharing a CDS stage and 4 analog memory banks with 4 capacitors each. Additional 2 capacitors are shared between the 4 pixels and are used for averaging (i.e. writing twice to the same memory capacitor) and binning operations. (B) waveform description of the basic in-pixel CDS and storage operation.**

After the operation described above is completed (in under 400  $\mu\text{sec}$ ), storage capacitors C1 in the entire array contain  $0^\circ$  data sampled from SG1, and respectively C3 capacitors contain  $180^\circ$  data sampled from SG2. The next step would be performing another exposure for sampling  $90^\circ$ - $270^\circ$ , and populating C2 and C4 with the corresponding data. Theoretically, at that point all the required data for calculating phase/depth exists. However, since each phase was collected from a specific storage gate, the data might not be perfectly symmetric. This is due to the fact that the two SGs may collect different amount of parasitic light, generate different amount of dark current, etc. which could lead to depth inaccuracy. In order to create symmetric data, we introduce the averaging mechanism: after storing phases  $0^\circ$  (from SG1),  $90^\circ$  (SG1),  $180^\circ$  (SG2), and  $270^\circ$  (SG2) in C1,2,3, and 4 respectively, we reverse the roles, and modulate with complementary phases:  $180^\circ$  -  $0^\circ$  and  $270^\circ$  -  $90^\circ$ .

Now, each storage gate stores the complementary phase to the one it stored previously, but we encounter a

problem: how to store this signal in a storage capacitor that already holds previous data without erasing it? This is done by temporarily storing the data on COUT1, and then connecting COUT1 to the relevant storage cap. The 2 capacitors are identical, and the new and previous signals are averaged through charge sharing. Eventually, each storage capacitor will hold signal of one phase ( $0^\circ$ ,  $90^\circ$ ,  $180^\circ$ , or  $270^\circ$ ), sampled and averaged equally from both storage gates, and therefore fully differential. The depth biases due to variations between storage gates will be eliminated from the data. The complete operation is summarized in Table 1.

Finally, the fact that every four vertically adjacent pixels share the same in-pixel CDS circuit and are connected to a set of 16 storage capacitors, allows store more than just 4 readouts, as long as the array is binned accordingly. This is very useful in order to resolve ambiguity issues due to phase wraparound that are inherent in iToF technology, which requires data from more than one frequency. This coincides nicely with the fact that longer ranges usually require pixel binning for improved SNR.

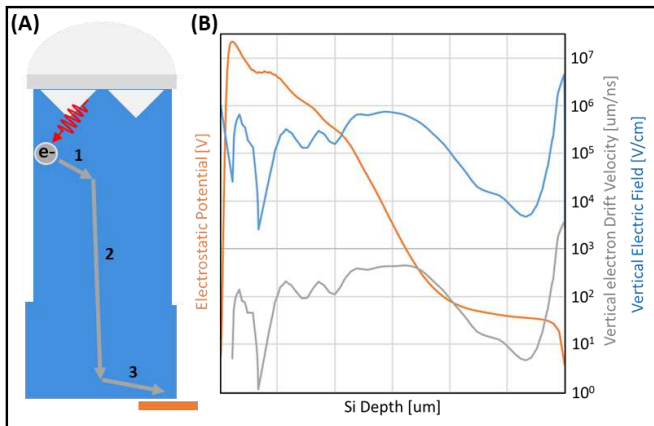
### Pinned Photodiode Optimization for High Modulation Contrast & Frequency

Achieving good depth quality in iToF sensors requires high modulation frequency, high modulation contrast, and high modulation uniformity between adjacent pixels and across the array. Those are achieved by design of the electric field in the photodiode, so that any photoelectron that is generated inside will quickly drift to the correct storage area. Photoelectrons that get stuck in areas with low electric field have chance of being integrated into the wrong storage, thus lowering the modulation contrast. Furthermore – the time it takes for the electron to cross the low-field area is in many cases sensitive to process variations and therefore might cause different modulation response in adjacent pixels that will translate into patterns in the depth image. Optimization the implant scheme and layout of the photodiode is required in order to eliminate those areas with potential pockets or low electric fields. In order to achieve maximal electron velocity, the photodiode was designed to guide the electron in the path described in Figure 3(A). After photogeneration the electron will drift toward the center of the photodiode, then drift toward the surface, and will finally be collected into the correct storage area under the effect of the modulation gates. A series of low-dose, high energy implants were introduced into the photodiode in order to

**Table 1- summary of the data contained in each storage capacitor after each exposure and in-pixel CDS cycle**

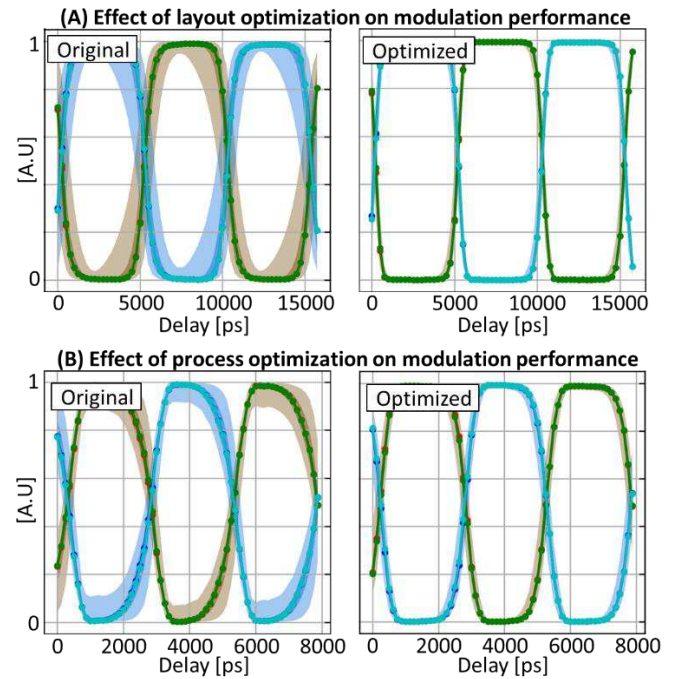
Stored data after Exposure #:	C1	C2	C3	C4
1	$\phi(0^\circ)_{SG1}$	–	$\phi(180^\circ)_{SG2}$	–
2	$\phi(0^\circ)_{SG1}$	$\phi(90^\circ)_{SG1}$	$\phi(180^\circ)_{SG2}$	$\phi(270^\circ)_{SG2}$
3	$[\phi(0^\circ)_{SG1} + \phi(0^\circ)_{SG2}]/2$	$\phi(90^\circ)_{SG1}$	$[\phi(180^\circ)_{SG1} + \phi(180^\circ)_{SG2}]/2$	$\phi(270^\circ)_{SG2}$
4 (Final)	$[\phi(0^\circ)_{SG1} + \phi(0^\circ)_{SG2}]/2$	$[\phi(90^\circ)_{SG1} + \phi(90^\circ)_{SG2}]/2$	$[\phi(180^\circ)_{SG1} + \phi(180^\circ)_{SG2}]/2$	$[\phi(270^\circ)_{SG1} + \phi(270^\circ)_{SG2}]/2$

achieve strong electric field and high electron drift velocity towards the surface. The implant scheme and photodiode layout were optimized using three dimensional TCAD simulations to make sure that an electron generated anywhere in the photodiode will be collected to the storage gate in under 1ns. Simulated cross-sections of the Electrostatic Potential, Electric Field, and electron Drift Velocity extracted from the center of the photodiode can be seen in figure 3B.



**Figure 3(A)** – a conceptual cross section of an iToF pixel photodiode, showing the path of the photoelectron. After photogeneration the electron quickly drifts towards the center of the photodiode (1), then towards the silicon surface (2), and finally is collected into the relevant storage gate (3). **(B)** – one-dimensional cross section generated using TCAD simulation showing the Electrostatic Potential, Electric Field, and Drift Velocity in the center of the photodiode.

As part of the pixel development an experiment was planned both around the layout and the implant recipe of the photodiode. The effect of the optimization on modulation performance is apparent in figure 4, which presents experimental results of scanning the pixel optical response vs. delay time using a pulsed laser with a very short (10s of picoseconds) pulse width. Each data point in the plot represents the average pixel response for a certain delay between the pixel modulation clock and the laser pulse. The dark shaded area (which is hardly visible) represents pixels within 1 standard deviation from the average modulation, and the light shaded area represents the 95<sup>th</sup> percentile of all pixels. Figure 4A shows the result of layout optimization, and specifically increasing the modulation gate (MG) length on the modulation uniformity. This measurement was performed in modulation frequency of 100MHz and with modulation voltage of 1.2V. It can be seen in the plot that the modulation shape and pixel to pixel variation improved significantly. Modulation contrast was not affected and remained at 96%.



**Figure 4** – Scans of the pixel response to fast modulation, measured with a short (<100 ps) pulse laser and a delay generator. Each data point represents the average pixel response per specific laser delay. **(A)** – scans with modulation frequency of 100MHz and modulation voltage of 1.2V. **(B)** – scans with modulation frequency of 200MHz and modulation voltage of 1.8V.

Figure 4B shows the result of implant dose optimization, and in this case specifically optimization of the implant dose of the deep and shallow n-type implants that form the photodiode. In this case, the measurement was performed in modulation frequency of 200MHz and modulation voltage of 1.8V, and the results show again significant improvement in modulation shape and pixel to pixel variations.

### Summary

We have presented a 1280x960 iToF sensor with a 3.5um pixel, in-pixel CDS circuit, ability to store 4 frames in voltage-domain storage capacitors, 38% QE @ 940nm, >25ke- full well capacity, and up to 200MHz demodulation frequency with >95% modulation contrast. Point cloud output from the sensor is presented in Figure 5. Part A of the figure shows a detailed depth scene captured in a single depth frame without temporal averaging. In Figure 5B the same point cloud is rotated about the Y-Axis, showing the high accuracy of the geometry and lack of flying pixels. Figure 6 presents a dynamic scene of a ball being thrown in the air with minimal motion artifacts, enabled by the in-pixel CDS and storage mechanism. Chip micrograph can be observed in Figure 7, and comparison to other recent iToF sensors in presented in Table 2.

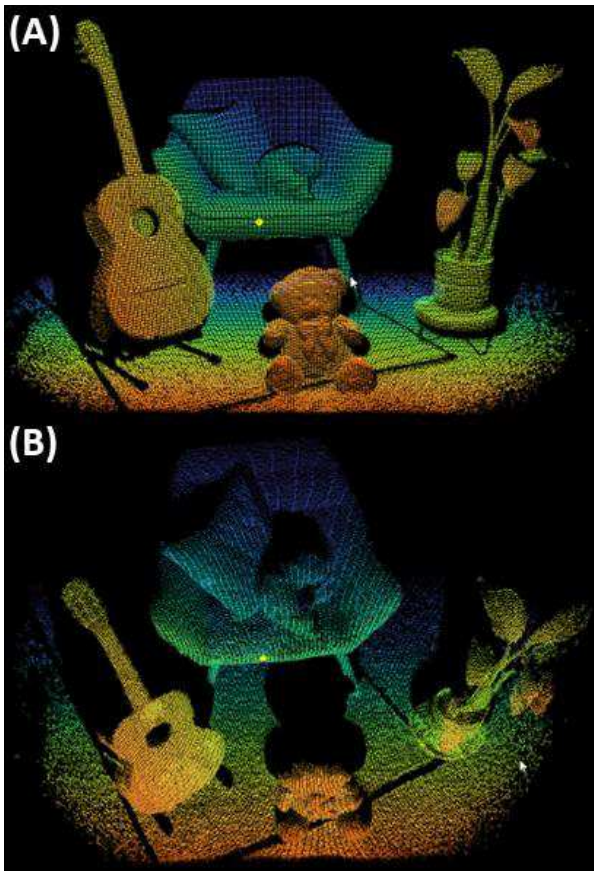


Figure 5 – (A) Point cloud from a detailed scene captured from a single depth frame. (B) same point cloud rotated about the Y-Axis.

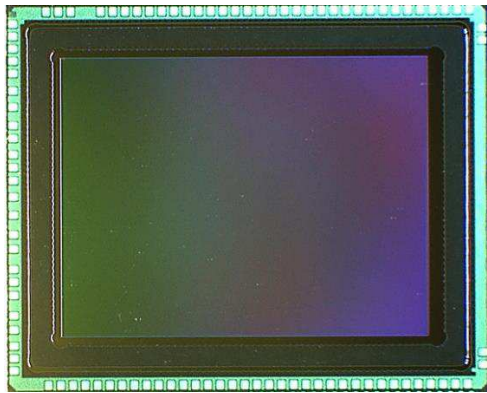


Figure 7 - Micrograph of the stacked die in CSP package. Chip dimensions are 6.1mm x 4.9mm.



Figure 6- A depth image of a ball being thrown in the air, showing valid depth readout from the moving ball, highlighting the contribution of the in-pixel storage architecture to reduced motion artifacts.

## References

- [1] R. Lange, "3D time-of-flight distance measurement with custom solid state image sensors in CMOS/CCD-technology," Ph.D. dissertation, UCLA, CA, USA, Univ. Siegen, Siegen, Germany, 2000.
- [2] Bamji, Cyrus, et al. "A Review of Indirect Time-of-Flight Technologies." IEEE Transactions on Electron Devices (2022).
- [3] Hatakeyama, Kunihiro, et al. "A Hybrid Indirect ToF Image Sensor for Long-Range 3D Depth Measurement under High Ambient Light Conditions." 2022 IEEE Symposium on VLSI Technology and Circuits (VLSI Technology and Circuits). IEEE, 2022.
- [4] Y. Ebiko et al., "Low power consumption and high resolution 1280×960 gate assisted photonic demodulator pixel for indirect time of flight," in IEDM Tech. Dig., Dec. 2020, p. 33
- [5] M.-S. Keel et al., "A 4-tap 3.5 μm 1.2 Mpixel indirect time-of flight CMOS image sensor with peak current mitigation and multi-user interference cancellation," in IEEE ISSCC Dig. Tech. Papers, Feb. 2021
- [6] M. Tsutsui et al. "A 3-Tap Global Shutter 5.0um Pixel with Background Canceling for 165MHz Modulated Pulsed Indirect Time-of-Flight Image Sensor" in IISW 2021.
- [7] Tubert, Cédric, et al. 4.6μm Low Power Indirect Time-of-Flight Pixel Achieving 88.5% Demodulation Contrast at 200MHz for 0.54 MPix Depth Camera." ESSDERC 2021-IEEE 51st European Solid-State Device Research Conference (ESSDERC). IEEE, 2021

Table 2-recent iToF sensor performance comparison

	This Work	IEDM 21' [4]	ISSCC 21'[5]	IISW21'[6]	ESSDERC21'[7]
Pixel Pitch	3.5um	3.5um	3.5um	5um	4.6um
Process	65nm / 65nm	90nm / 65nm	65nm / 65nm	65nm / 65nm	65nm / 40nm
Sensor Resolution	1280 x 960	1280 x 960	1280 x 960	640 x 480	672 x 804
Max. Mod. frequency	200MHz	120MHz	200MHz	165MHz	200MHz
Modulation Contrast	96% @ 100MHz 95% @ 200MHz	96% @ 100MHz	96% @ 100MHz 80% @ 200MHz	88% @ 100MHz 81% @ 165MHz	88.5@200MHz
Read noise (direct readout)	3.5e-	-	3.4e-	3.5e-	4.3e-
Read noise (storage caps)	7.0e-	-	-	-	-
Linear Full Well per tap	25ke-	18ke-	-	11ke-	-
940nm QE	38%	32%	38%	21%	18.5%



# A 320×232 LiDAR Sensor with 24dB Time-Amplified and Phase-Revolved TDC

Chin Yin, Shang-Fu Yeh, Chiao-Yi Huang, Hon-Yih Tu, Meng-Hsiu Wu, Tzu-Jui Wang, Kuo-Chin Huang and Calvin Yi-Ping Chao  
 Taiwan Semiconductor Manufacturing Company, Hsinchu, Taiwan, ROC  
 Tel: (886) 3-5636688 ext 797-8127, email: cyin@tsmc.com

**Abstract**—This paper presents a 320×232, 6.84μm SPAD 3D-stacked BSI LiDAR sensor. With a 24dB Time Amplifier pre-amplifying the Time-of-Flight and a TDC performing phase-revolved conversion, a 3.81ps TDC resolution with [-0.3, 0.4] DNLs is verified. By utilizing time-correlated single-photon counting with the proposed TDC, a 4-bit data compression is demonstrated without sacrificing the image quality. The prototype depth imager achieves 0.5cm distance accuracy and 24 frames/s ToF image rate.

**Keywords**—SPAD, LiDAR, ToF, TDC, Time Amplifier

## I. INTRODUCTION

Depth imaging using light detection and ranging (LiDAR) technology is a key feature in many emerging applications, e.g., autonomous driving, industrial modeling, and interactive AR/VR systems. Direct time-of-flight (D-ToF) method can achieve long-range detection and high frame rates. With powerful pulsed laser and 3D-stacked back-side illumination (BSI) single-photon avalanche diode (SPAD) technology, low system jitter contributed from optics and silicon is achieved. To further realize high depth resolution with large image format, a highly accurate, parallel time-to-digital converter (TDC) is the key component. Flash systems adopt pixel-parallel or group-parallel TDC to reach high frame rates, sacrificing the pixel pitch [1-4], TDC dynamic range and uniformity control. Column-parallel TDC configuration is suitable for narrow pixel pitch, uniform readout quality and practical data throughput.

The flash TDC structure receives multiple time-resolved clock phases from a global delay-locked loop (DLL) block. However, the systematic clock skew decreases the TDC linearity, and the time resolution is limited by the DLL frequency. The conventional D-ToF sensor targets several-centimeters distance accuracy. Time-correlated single photon counting (TCSPC) is utilized to suppress the ambient light interference and jitter distribution, at the cost of huge data oversampling and histogramming. There is a trade-off between TDC dynamic range and the subsequent data processing effort.

In this paper, we propose a column-parallel 24dB time-amplified and phase-revolved (PR) TDC structure optimized

for sub-centimeter distance accuracy and data processing reduction. Fig. 1 shows the conceptual architecture, composed of an optical module with uniformly diffused pulsed laser and near Infrared (NIR) lens, a SPAD detector, a front-end time-amplifier (TA) with 24dB gain, and a latch-based TDC circuit. The TDC receives multiple clock phases from the global DLL, and the phase orders are revolved per TDC conversion. Because of the PR multiplexer, we implemented two modes optimized for linearity boost and data compressive.

## II. OPERATION MODES AND PRINCIPLE

### A. Linearity Boost Mode

In the linearity boost mode (Fig. 2), all the latch cells in TDC circuit are enabled. The latches sample the monotonic phases from DLL, and reconstruct the latched thermometer codes to binary digits. The intrinsic clock skews from M-stage DLL delay cells cause the differential non-linearity (DNL) of TDC. Then the DLL phase orders are revolved one step in the next TDC conversion, results in one digit shift of DNLs. The DNLs form a fixed pattern cycle due to the nature of DLL that locks the clock in with the clock out of the delay chain. Therefore, after M steps revolution, the DNL cycle finishes a full round of shifts. Because of the oversampling feature of TCSPC, the system collects M multiple times of TDC codes, and the DNLs are self-calibrated to zero after histogram testing theoretically.

### B. Data Compressive Mode

In the data compressive mode (Fig. 3a), only one latch cell is enabled. The TDC latches only one phase as least significant bits (LSBs) part, and truncate the lower  $\log_2 M$  bits. Then the DLL phase orders are revolved one step at next TDC conversion, which results in TDC intervals shifting by one phase offset. After M multiple times of TDC codes histogramming, a simple average process reconstructs the histogram peak as nominal TCSPC. The data compressive mode effectively reduces the data throughput without TDC LSB codes, and the pre-average process suppresses the front-end jitter represented as full-width half maximum (FWHM) of the histogram (Fig. 3b). The suppression trend follows the oversampling theory. By choosing M=16 and adopting an

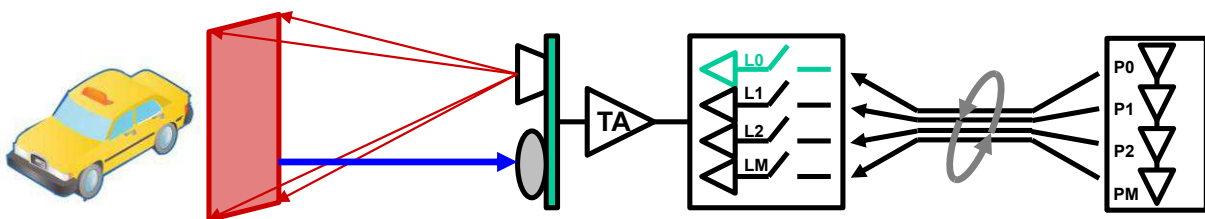


Figure 1. Concept of proposed LiDAR sensor.

averaging process, 4 bits data length is truncated and the FWHM is reduced by 4 times.

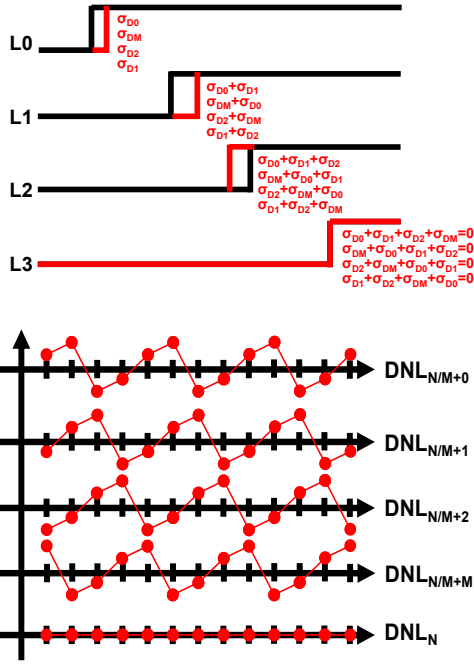


Figure 2. Linearity boost modes

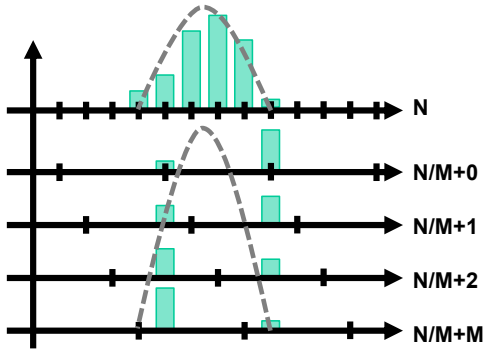


Figure 3a. Data compressive mode

### III. PROPOSED LIDAR SENSOR CIRCUITS

Fig. 4 shows the system block and the key circuit component diagrams of this LiDAR sensor. The detector array includes  $320 \times 232$   $6.84\mu\text{m}$  SPADs fabricated in 45nm BSI CIS node. With 3D stacked technology, the top layer SPADs are pixel-wise bonded with the bottom layer of active quenching and re-charging (AQRC) circuits in 22nm logic process. The peripheral parts include a column-parallel digital timer cooperated with AQRC pixel for SPAD hold-time control, a row selector for D-ToF line scanning, a column TA to extend the succeeding TDC dynamic range, a column PR TDC and counter receiving the multiplexed clock phases from global PLL and DLL, the column serializer controlled by APR processor, and a low-voltage differential signaling (LVDS) interface transferring data for off-chip processing.

#### A. Active Quenching and Re-Charging

In the pixel-parallel AQRC (Fig. 5), the effective quench resistance is controlled by a quench bias voltage, and the hold-time is controlled by either the internal switch-capacitance integrator or the column AQRC timer.

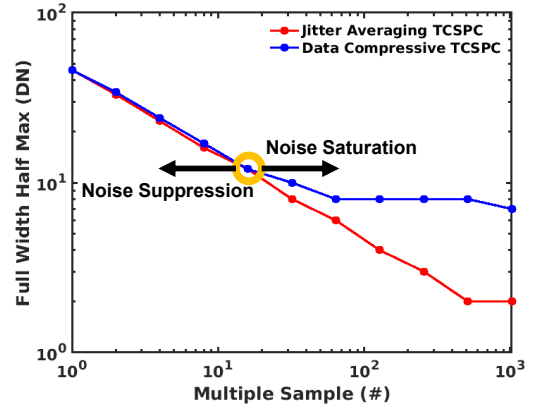


Figure 3b. Data compressive mode jitter

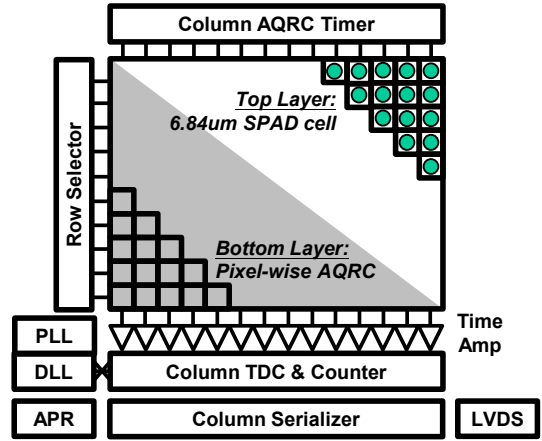


Figure 4. System block diagram

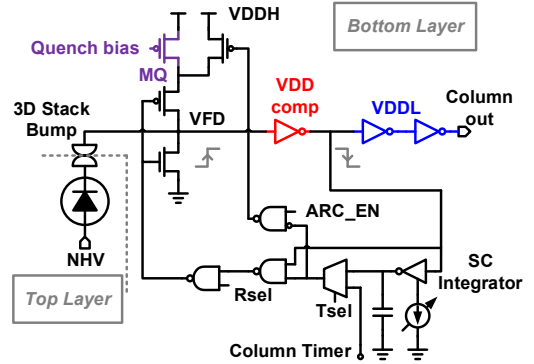


Figure 5. Pixel-parallel AQRC

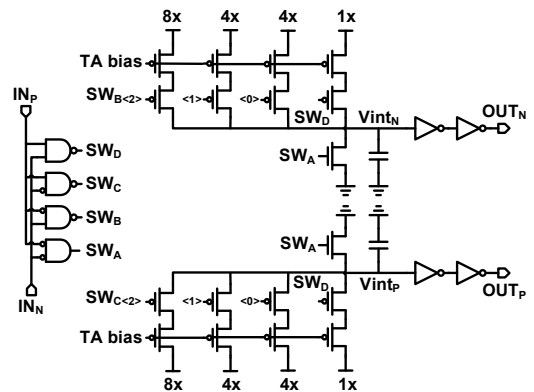


Figure 6. Time amplifier

## B. Time Amplifier

At the initial state of time amplifier (Fig. 6), both  $V_{intP}$  and  $V_{intN}$  are reset to low. Once the  $IN_N$  receives a rising edge first, the  $SW_B$  turn on and  $V_{intN}$  start to be integrated by a large current. After certain time-of-flight interval,  $IN_P$  is triggered to high, controlling both  $V_{intP}$  and  $V_{intN}$  to be integrated by a small current ( $SW_D$ ).  $V_{intN}$  reaches the inverter threshold first followed by  $V_{intP}$ , and the time amplification ratio between  $OUT_{P/N}$  and  $IN_{P/N}$  is proportional to the current ratio between  $SW_B$  and  $SW_D$ . We choose  $4\times$ ,  $8\times$ ,  $16\times$  gain ratios in this design.

## C. Phase-Revolved TDC

In the proposed PR TDC circuit (Fig. 7), 1GHz differential clocks  $CK_{INP}$  and  $CK_{INN}$  are delayed and locked by the 8-stage global DLL block, providing total 16 equivalent delayed phases. The 16 phases are multiplexed to each column TDC circuit, which includes 16 latch cells as LSBs part and 8-bit ripple counters as MSBs part. The latches are event-driven by the preceding SPAD event, and the latched thermometer codes are combined with MSB parts to form a 12-bit TDC output. Two identical TDC circuits per column are used to achieve digital correlated double sampling (DCDS) purpose. The multiplexed 16 phase orders are revolved according to the frame index signal, to realize either linearity boost or data compressive functions.

## D. Operation Timing diagram

Fig. 8 is the timing diagram of the proposed LiDAR sensor with TA and PR TDC. The 232 lines are scanned sequentially into one frame. In one line period, a LASER pulse is emitted and reflected, triggering the SPAD to avalanche at  $V_{FD}$ . The  $V_{FD}$  falling edge enable the hold time control, which limits only one pulse within TDC window. A reference LASER starts ( $IN_N$ ) and the SPAD column-out ( $IN_P$ ) inputs to the TA, and the time interval is amplified by up to  $16\times$  times. The amplified signals ( $OUT_{P/N}$ ) are sent to the TDC. Two input signals are converted by dual TDC circuits separately, and processes the DCDS result. In mode 1, all the latches are enabled, while in mode 2, only the first latch is activated. The DLL outputted phase orders are revolved with frame index, effectively flattening the clock skew. The TDC 12-bit result is then stored into line buffer, and output through LVDS driver serially at next line time.

## IV. MEASUREMENT AND SPECIFICATION

Fig. 9 presents the measurement setup. The LiDAR system applies 940nm 100W 20° beam angle pulsed VCSEL diode and the sensor part is assembled with F1.4 8mm lens and 940nm BPF. The in-lab targets under test and environment are limited to a range of a few meters.

### A. Linearity Boost Mode

Fig. 10 shows the characterization result. After the proposed PR self-calibration, the DNLs are improved from  $[-0.9, 0.9]$  to  $[-0.3, 0.4]$ . The clock-induced skews are flattened and the DNLs are only limited by the clock tree routing mismatch. The transfer curves of the proposed TDC without and with TA in  $4\times$ ,  $8\times$ ,  $16\times$  gains are measured. The gain slopes follow the TA current design, with saturation levels slightly shifted by switch coupling effect. In the ground-truth distance measurement, 0.5cm depth accuracy is achieved across 100 centimeters measured range.

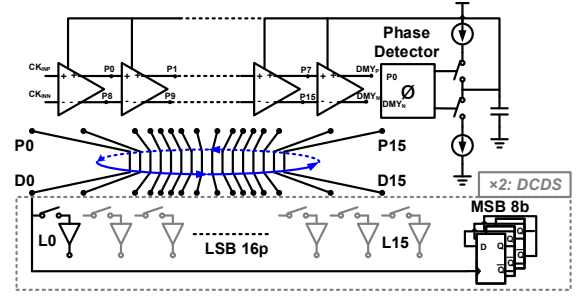


Figure 7. Phase-revolved TDC

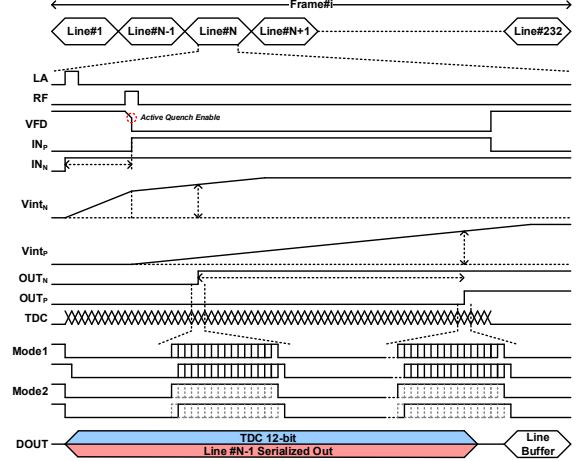


Figure 8. Timing diagram

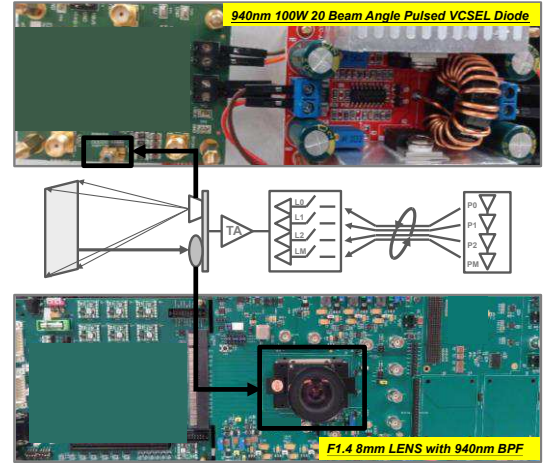


Figure 9. Measurement setup

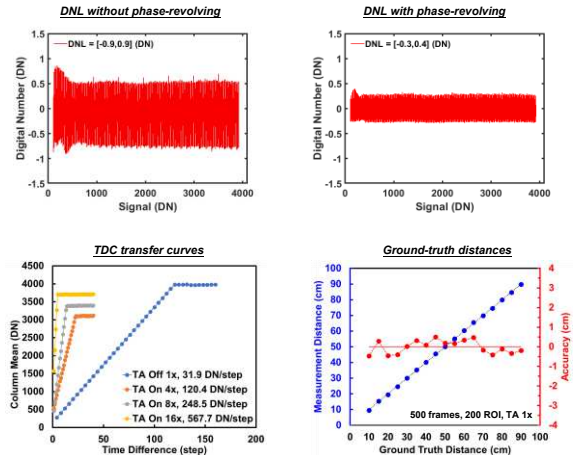


Figure 10. TDC characterization, Time-amp linearity, ground-truth

## B. Data Compressive Mode

Fig. 11 collects 10000 frames for depth image and histogramming demonstration. Comparing to nominal TCSPC depth image, the PR TDC reconstruct the 16× (4-bit) data compression depth image without sacrificing the image quality.

## C. Performance

The performance compared with state-of-the-arts are listed in Fig. 12. The 6.84μm pitch 320×232 LiDAR sensor reaches a 3.81ps TDC resolution with [-0.3, 0.4] DNLs by proposed time-amplified and phase-revolved functions. Fig. 13 shows the micrograph of this 3D stacking LiDAR sensor and the performance summary, a total 96dB dynamic range, a 0.5cm distance accuracy and 24 frames/s ToF image rate are achieved.

## V. CONCLUSION

This 320×232, 6.84μm SPAD 3D-stacked BSI LiDAR sensor integrates AQRC circuit, 24dB Time Amplifier and a 3.81ps resolution with [-0.3, 0.4] DNLs TDC for TCSPC operation. The 4-bit data compression depth image is demonstrated, and the 0.5cm distance accuracy with 24 frames/s ToF image rate are measured. The 3D depth model presents sub-centimeter depth resolution (Fig. 14).

## ACKNOWLEDGMENT

The authors would like to thank the process team for the design of the SPAD device and the fabrication of the test chip.

## REFERENCES

- [1] 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", ISSCC Dig. Tech. Papers, pp. 112-113, Feb. 2021.
- [2] O. Kumagai, et al., "A 189×600 Back-Illuminated Stacked SPAD Direct Time-of-Flight Depth Sensor for Automotive LiDAR Systems", ISSCC Dig. Tech. Papers, pp. 110-111, Feb. 2021.
- [3] S. Park, et al., "An 80×60 Flash LiDAR Sensor with In-Pixel Histogramming TDC Based on Quaternary Search and Time-Gated Δ-Intensity Phase Detection for 45m Detectable Range and Background Light Cancellation", ISSCC Dig. Tech. Papers, pp. 98-99, Feb. 2022.
- [4] 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", ISSCC Dig. Tech. Papers, pp. 96-97, Feb. 2022.

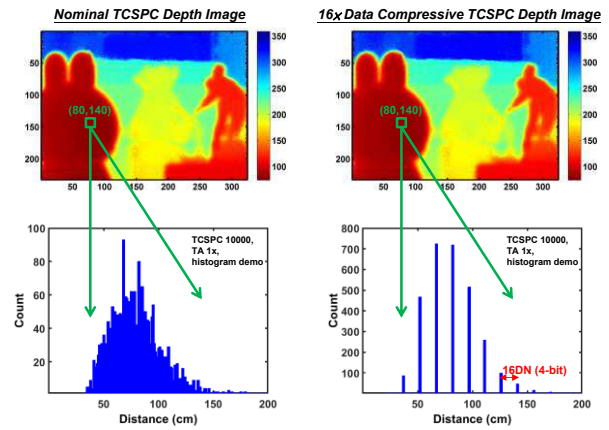


Figure 11. Demo of of TCSPC depth images wo/wi data compression

Parameter	Unit	This Work	[1]	[2]	[3]	[4]
Technology	nm	45/22 SPAD CMOS	45/22 SPAD CMOS	90/40 SPAD CMOS	110	110
TOF format (H×V)	-	320×232	256×128	168×63	80×60	64×64
TOF method	-	Direct	Direct	Direct	Hybrid	Direct
Pixel pitch	μm	6.84	7	10	75	48
SPAD structure	-	AQ+ARC	Coin.	Coin.	Hist	Coin.
TDC architecture	-	Column	256-to-1	9-to-1	6-to-1	Pixel TDC
TDC depth	bit	12 & TimeAmp	14	12	13	16
TDC resolution	ps	Intrinsic: 61 TA 4×: 15.25 TA 8×: 7.63 TA 16×: 3.81	60	1000	100	250
TDC linearity (DNL)	DN	+0.4/-0.3	±0.05***	N/A	N/A	+0.79/-0.61
LASER projection	-	Flash	Flash	Flash	Flash	Flash
Wavelength	nm	940	780	905	905	905
Distance range	m	256/10*	100	200/150	45	8.2
Distance accuracy	cm	0.5	7	15-30	2.5	15
Repetition rate	Hz	2400	500k	N/A	100k	6250
TOF image rate	fps	24**	N/A	20	30	25

\* Optical limitation  
\*\* 100 frames TOF  
\*\*\* After Calibration

Figure 12. Comparison table

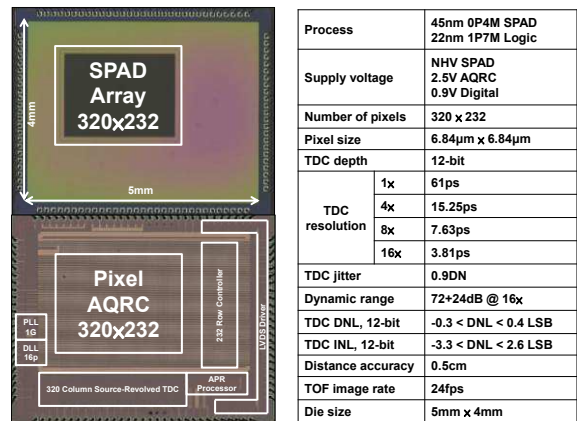


Figure 13. Chip micrograph and summary

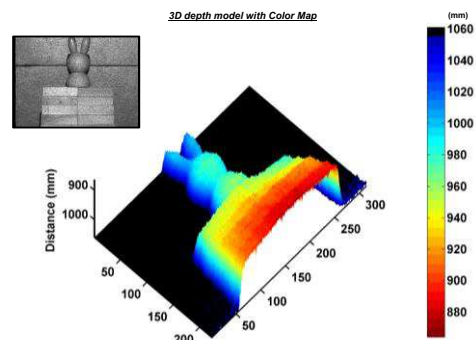


Figure 14. 3D depth model

# A 648 x 484-Pixel 4-Tap Hybrid Time-of-Flight Image Sensor with 8 and 12 Phase Demodulation for Long-Range Indoor and Outdoor Operations

Kamel Mars<sup>1</sup>, Kensuke Sakai<sup>1</sup>, Yugo Nakatani<sup>1</sup>, Masashi Hakamata<sup>1</sup>, Keita Yasutomi<sup>1</sup>, De Xing Lioe<sup>1</sup>, Keiichiro Kagawa<sup>1</sup>, Tomoyuki Akahori<sup>2</sup>, Tomohiko Kosugi<sup>2</sup>, Satoshi Aoyama<sup>2</sup>, Shoji Kawahito<sup>1,2</sup>

<sup>1</sup> : Research Institute of Electronics, Shizuoka University, Hamamatsu, 432-8011, Japan

<sup>2</sup> : Brookman Technology, Inc., Hamamatsu, 430-0936, Japan

email: [kamel@idl.rie.shizuoka.ac.jp](mailto:kamel@idl.rie.shizuoka.ac.jp) TEL: +81-53-478-1341 FAX: +81-53-478-3251

## I. INTRODUCTION

For 3D depth sensing, direct ToF (dToF) sensors are becoming popular particularly for long-range outdoor applications, but the dToFs has a difficulty of high depth-image resolution if high ambient light operation is required [1]-[3]. On the other hand, indirect TOF (iTOF) image sensors using Continuous-Wave (CW) demodulation pixels are good solution for high image resolution ToF cameras [4]-[7], but the iToFs have a difficulty of long-range measurements under strong ambient light because the iToF pixels using 50% duty-cycle operation suffer from strong ambient light. This paper presents a 100klux ambient light tolerant VGA-resolution hybrid-type ToF image sensor, which is basically a short-pulse (SP) based iToF but practically uses a concept of the dToF for finding the received light-pulse location as a digital number in the multiple time windows prepared by multi-tap pixel outputs and multiple subframes [8-9]. The presented hybrid ToF technique has distinct advantages over the CW-iTOF counterpart because of the reduced in-pixel ambient-light charge acquisition and the resulting small amount of ambient-light shot noise for outdoor operation.

## II. SENSOR ARCHITECTURE AND OPERATION

Figure 1 shows the sensor chip architecture and 4-tap demodulation pixel with a lateral-drift pinned-photo diode structure and lateral electric-field control gates, G1~G4 and drain gate GD. Fast driving of MOS gates can be achieved using a shared in-pixel driver with two driver circuits (i.e. Top and bottom driver) where each of them is used for one half of the pixel array for better drivability. In a pixel pitch of 16.8 $\mu$ m, 4PGAs (4.2 $\mu$ m pitch) and 4 succeeding 12b cyclic A/D converter

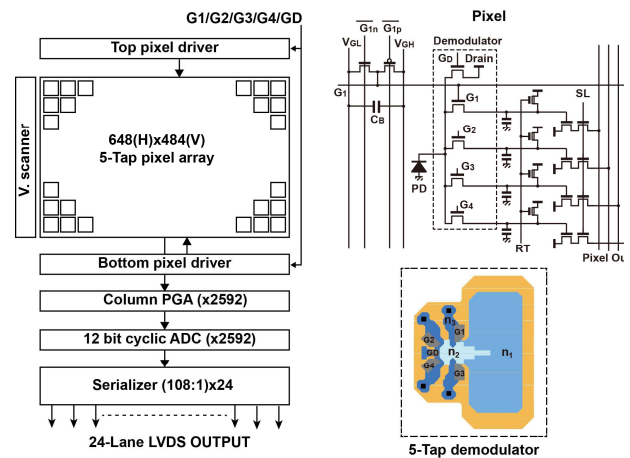


Figure 1: Prototype chip photograph (Left), Pixel circuit and demodulator diagram (Right).

(4.2 $\mu$ m pitch) are arranged at the column allowing a fully-parallel readout of 4-taps per pixel. With 24 serializers (108-to-1), and 24-lane LVDS outputs, the TOF image data are readout at maximally 11.5Gbps ensuring a fast readout time of the full 4-tap VGA images in just 1.45ms allowing a higher frame rate to be achieved using hybrid-type ToF measurements.

In order to exploit the feature of the SP modulation method while attaining wide distance measurement range, the proposed method uses multiple time-gating pulses and range-shifted TOF measurements with multiple sub-frame (SF) readouts as shown in Figure 2 and Figure 3 for both indoor and outdoor operation respectively. Outdoor operation as depicted by the timing diagram in Figure 3, uses four subframes, where the first two are used for the near depth zone and the other two are used for far depth zone. By using 8 gating pulses, a measurable range of 0 to 7TP (e. g., 11.8m for TP of 11.25ns, where TP is the light-pulse width) is

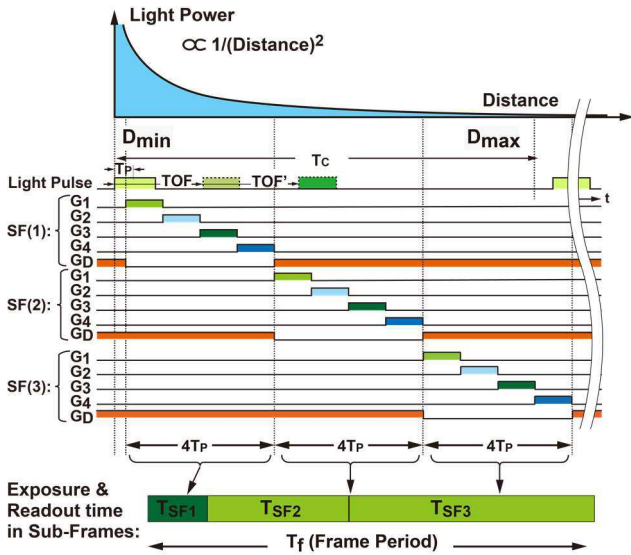


Figure 2: 3-subframes gates timing chart for indoor measurement

attained by a careful selection of the exposure time ratio based on the fact that light power attenuation is inversely proportional to the square of the distance. Two sub-frame readouts with switching gate pulses between G1 and G3, or G2 and G4 for each depth zone as shown in Figure 3 are used with the aim of reducing ambient light-induced depth offset when a tap-to-tap conversion gain deviation is considered. By calculating the difference of two signals of SF(1) to SF(2) and SF(3) to SF(4), signal components are added while the residual ambient-light offset components due to the tap-to-tap gain deviation are canceled out.

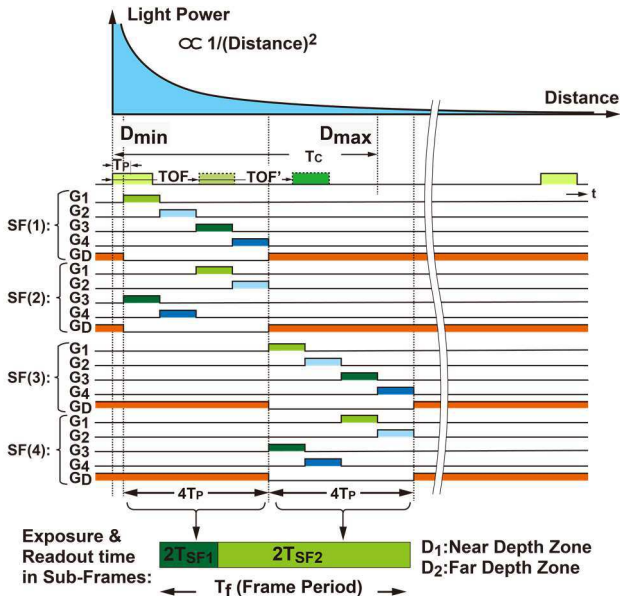


Figure 3: 4-subframes gates timing chart for outdoor measurement

The output signal of  $i$ th tap,  $S_i$  if the conversion gain is deviated from tap to tap can be expressed as

$$S_i = (G + \Delta G_i)(X_{Sk} + X_{Ak}) \quad (1)$$

where  $X_{Sk}$  is the signal charge generated at  $k$ th gate timing,  $X_{Ak}$  is the ambient light charge generated at  $k$ th gate timing,  $G$  is the conversion gain factor and  $\Delta G_i$  is the deviation of  $G$ . Based on the timing chart of the outdoor operation, the signal differences in the first subframe,  $S_{13} = S_1 - S_3$  and  $S_{24} = S_2 - S_4$  can be expressed as

$$S_{13}^{SF} = (G + \Delta G_1)(X_{S1}^{SF1} + X_{A1}^{SF1}) - (G + \Delta G_3)(X_{S3}^{SF1} + X_{A3}^{SF1}) \quad (2)$$

and

$$S_{24}^{SF1} = (G + \Delta G_2)(X_{S2}^{SF1} + X_{A2}^{SF1}) - (G + \Delta G_4)(X_{S4}^{SF1} + X_{A4}^{SF1}) \quad (3)$$

Since the order of tap opening is being inverted in the second subframe,  $S_{13} = S_1 - S_3$  and  $S_{24} = S_2 - S_4$  can be expressed as

$$S_{13}^{SF2} = (G + \Delta G_1)(X_{S3}^{SF2} + X_{A3}^{SF2}) - (G + \Delta G_3)(X_{S1}^{SF2} + X_{A1}^{SF2}) \quad (4)$$

and

$$S_{24}^{SF2} = (G + \Delta G_2)(X_{S4}^{SF2} + X_{A4}^{SF2}) - (G + \Delta G_4)(X_{S2}^{SF2} + X_{A2}^{SF2}) \quad (5)$$

Since gates opening time is set to be the same, and careful exposure time setting and fast readout time are performed, we can assume that in each subframe, the ambient light components of each tap are equal, i.e.,  $X_{A1}^{SF} = X_{A3}^{SF} \equiv X_A^{SF}$  and  $X_{A1}^{SF2} = X_{A3}^{SF2} \equiv X_A^{SF2}$ . Equations (2) and (4) can be simplified by the following two equations:

$$S_{13}^{SF1} = G(X_{S1}^{SF1} - X_{S3}^{SF1}) + \Delta G_1 X_{S1}^{SF1} - \Delta G_1 X_{S3}^{SF1} + (\Delta G_1 - \Delta G_3) X_A^{SF1} \quad (6)$$

and

$$S_{13}^{SF2} = G(X_{S3}^{SF2} - X_{S1}^{SF2}) + \Delta G_1 X_{S3}^{SF2} - \Delta G_1 X_{S1}^{SF2} + (\Delta G_1 - \Delta G_3) X_A^{SF2} \quad (7)$$

The deviation factor due to ambient light in the first and second subframe  $\Delta S_A^{SF1}$  and  $\Delta S_A^{SF2}$  can be expressed as follow:

$$\Delta S_A^{SF1} = (\Delta G_1 - \Delta G_3) X_A^{SF1} \quad (8)$$

and

$$\Delta S_A^{SF2} = (\Delta G_1 - \Delta G_3) X_A^{SF2} \quad (9)$$

By taking the difference of two signal component of the first and second subframe ( $S_{13} = S_{13}^{SF1} - S_{13}^{SF2}$ ), the deviation factor can be express as

$$\Delta S_A^{SF12} = (\Delta G_1 - \Delta G_3)(X_A^{SF1} - X_A^{SF2}) \quad (10)$$

Equations 6 to 10 show that the signal component can be doubled while the ambient light component is canceled out and taps deviation is being minimized by calculating the difference of two subframes signals if background scene is not largely changed during the two subframes.

The same procedure is also used for tap 2 and 4 and performed for the far range (i.e. subframes 3 and 4) too.

For indoor operations where the background ambient light is not strong, 3 SFs are used as shown by the timing chart depicted in Figure 2 and 12 successive gating pulses can be used for 20m-range measurements of 0 to 11TP (i.e. 18.56m for TP of 11.25ns).

### III. MEASUREMENT RESULTS

In order to measure the proposed design, a prototype chip where the photomicrograph is shown in Figure 4 has been fabricated using 0.11 $\mu$ m COMS image sensor process. Table 1 shows a summary of specifications and basic characteristics of the chip. Figure 5 shows the measured response curve of the 3-subframes (Top) and 4-subframes (bottom) time-gating to the delay of light pulse. In 4-subframes case intended to be used for outdoor operation, only difference of two signals of SF(1) to SF(2) and SF(3) to SF(4) is depicted.

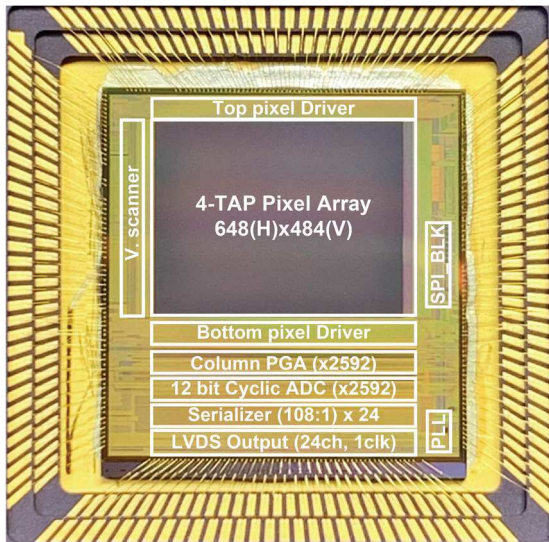


Figure (4): Prototype chip micro-photograph

Table 1: Summary of the prototype CMOS imager performance

Parameter	value
Technology	0.11 $\mu$ m CMOS Image sensor
Number of Pixels	648 (H) $\times$ 480 (V)
Pixel size	16.8 $\mu$ m $\times$ 16.8 $\mu$ m
Chip size	14.92 mm $\times$ 15.5 mm
ADC resolution	12 bits cyclic
Readout time	1.45 ms (12 bit)
Conversion gain	10.0 $\mu$ V/e <sup>-</sup>
Quantum Efficiency	18.6 % @ (940 nm)

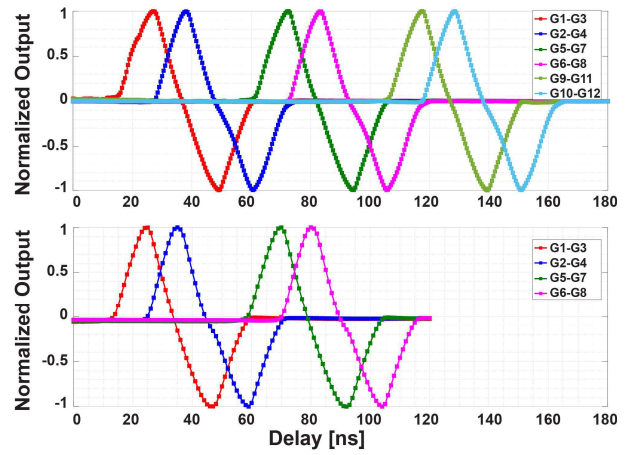


Figure 5: Pixel response to the light-pulse delay: 3-zone curves (top) and 2-zone curves (bottom).

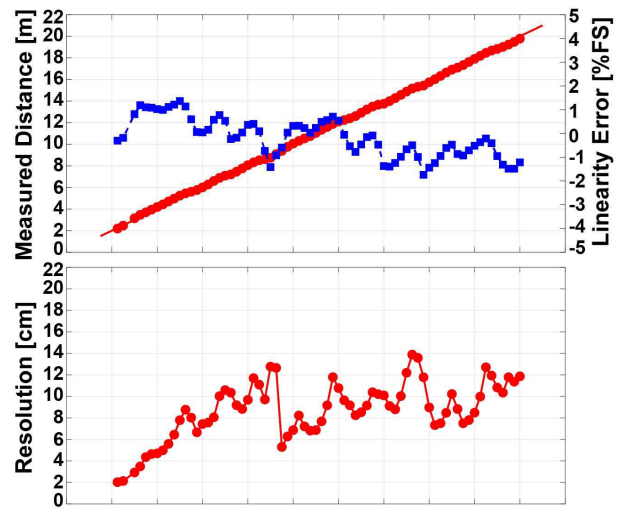


Figure 6: Indoor Distance, linearity and depth-noise measurements.

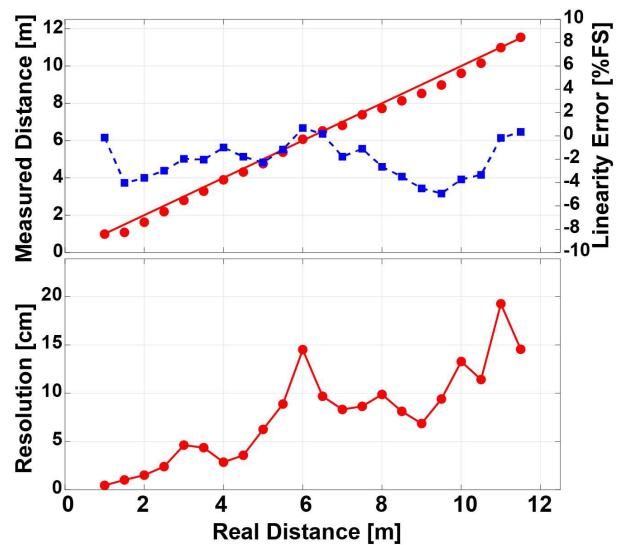


Figure 7: Outdoor distance, linearity and depth-noise measurement under 100klx

Figure 6 and Figure 7 shows the measured distance, linearity and depth noise (precision) for indoor (2-20m), and outdoor (1 to 11.5m) respectively. In the indoor measurements as depicted in Figure 6, the maximum non-linearity is 2%, and the maximum depth noise is 1.7% in the entire distance range. At 20m, a depth noise of 0.6% is attained. On the other hand, in the outdoor measurements as depicted in Figure 7, the maximum non-linearity is 4% and the depth noise at 10m is 1.6%. For both indoor and outdoor operation, the light pulse width and gates pulse with was set to 11ns and sensitivity correction between taps has been performed. Figure 8 shows the captured depth images for indoor (20m) and outdoor (10m, 110klux) measurements.

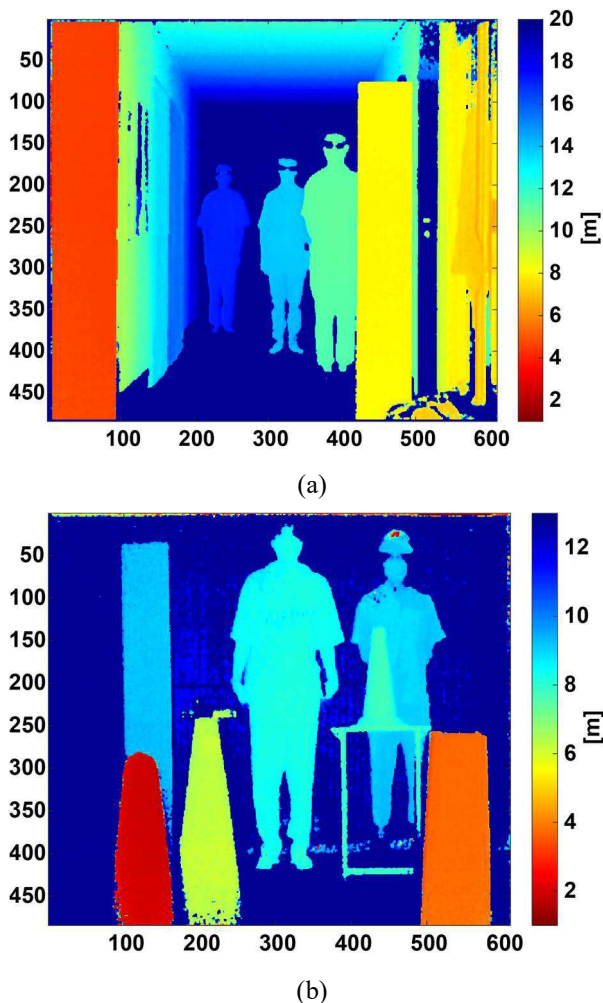


Figure 8: Captured TOF Images: (a) Indoor, (b) outdoor

#### ACKNOWLEDGMENT

This work was partly supported by Adaptable and Seamless Technology Transfer Program through Target-driven R&D (A-STEP) from Japan Science and Technology Agency (JST) Grant Number JPMJTR211A. It was also partly supported by A JSPS

KAKENHI Grand Number 18H05240, 19H02194, and the Center of Innovation Program. The authors also appreciate DB HiTek for the prototype chip fabrication.

#### REFERENCES

- [1] H. Seo et al., "Direct TOF Scanning LiDAR Sensor With Two-Step Multievent Histogramming TDC and Embedded Interference Filter," in *IEEE Journal of Solid-State Circuits*, vol. 56, no. 4, pp. 1022-1035, April 2021.
- [2] B. Kim, S. Park, J. -H. Chun, J. Choi and S. -J. Kim, "7.2 A 48×40 13.5mm Depth Resolution Flash LiDAR Sensor with In-Pixel Zoom Histogramming Time-to-Digital Converter," in *IEEE ISSCC Dig. Tech. Papers*, Feb. 2021, pp. 108–110.
- [3] O. Kumagai et al., "A 189 × 600 back-illuminated stacked SPAD direct time-of-flight depth sensor for automotive LiDAR systems," in *IEEE ISSCC Dig. Tech. Papers*, Feb. 2021, pp. 110–111.
- [4] C. S. Bamji et al., "1Mpixel 65nm 320MHz demodulated TOF image sensor with 3.5um global shutter pixels and analog binning," in *IEEE ISSCC Dig. Tech. Papers*, Feb. 2018, pp. 94–95.
- [5] M.-S. Keel et al., "A 4-tap 3.5μm 1.2Mpixel indirect time-of- flight CMOS image sensor with peak current mitigation and multi-user interference cancellation," in *IEEE ISSCC Dig. Tech. Papers*, Feb. 2021, pp. 105–106.
- [6] D. Kim et al., "Indirect time-of-flight CMOS image sensor with on-chip background light cancelling and pseudo-four-tap/two-tap hybrid imaging for motion artifact suppression," *IEEE J. Solid-State Circuit*, vol. 55, no. 11, pp. 2849–2865, Nov. 2020.
- [7] Y. Ebiko et al., "Low power consumption and high resolution 1280x960 gate assisted photonic demodulator pixel for indirect time of flight," in *Int. Electron Device Meeting Tech. Dig.*, Dec. 2020, pp. 721–724.
- [8] K. Hatakeyama et al., "A hybrid indirect ToF image sensor for long-range 3D depth measurement under high ambient light conditions," in *Proc. IEEE Symp. VLSI Technol. Circuits*, Jun. 2022, pp. 46–47
- [9] S. Kawahito, K. Yasutomi and K. Mars, "Hybrid Time-of-Flight Image Sensors for Middle-Range Outdoor Applications," in *IEEE Open Journal of the Solid-State Circuits Society*, vol. 2, pp. 38-49, 2022



# Tap mismatch mitigation of 3 $\mu\text{m}$ 2-tap pixels of indirect Time-of-Flight image sensor for high-speed depth mapping

Yuhi Yorikado<sup>1</sup>, Sozo Yokogawa<sup>1</sup>, Chihiro Okada<sup>1</sup>, Komomo Kodama<sup>1</sup>, Risa Iwashita<sup>1</sup>, Katsumi Honda<sup>1</sup>, Takahiro Hamasaki<sup>1</sup>, Yuki Hanabusa<sup>1</sup>, Shohei Yoshitsune<sup>2</sup>, Kei Nagoya<sup>2</sup>, Masatsugu Desaki<sup>2</sup>, Shota Hida<sup>2</sup>, Hayato Wakabayashi<sup>1</sup>, Fumihiko Koga<sup>1</sup>

1: Sony Semiconductor Solutions Corporation, 2: Sony Semiconductor Manufacturing Corporation  
4-14-1 Asahi-cho, Atsugi, Kanagawa, Japan, +81-50-3141-3782, [Yuhi.Yorikado@sony.com](mailto:Yuhi.Yorikado@sony.com)

**Abstract** This paper presents the development of a VGA-resolution stacked back-illuminated (BI) indirect time-of-flight (iToF) image sensor with 3.0  $\mu\text{m}$  2-tap pixels. Key features of the iToF image sensor include a quantum efficiency (QE) of 38% at 940 nm, a full well capacity (FWC) of 37 ke-, demodulation contrast (Cmod) of 88% at 200 MHz, and parasitic light sensitivity (PLS) mismatch of less than -50 dB across the entire image area. Additionally, a novel 2-frame sequence without anti-frames was found to maintain comparable depth noise performance with the 4-frame sequence in both indoor and outdoor conditions. These characteristics make the sensor suitable for low-power, high depth frame rate 3D imaging in a variety of applications.

## I. Introduction

3D sensing technologies have become increasingly important for a wide range of applications, including LiDAR for automotive applications, AR/VR for HMD and metaverse applications. One promising 3D sensing technology is the iToF image sensor, which offers easy access to high-resolution 3D mapping. Although the potential applications for iToF image sensors are numerous, there is room for improvement in the technology. In general, iToF image sensors require four-phase data ( $0^\circ$ ,  $180^\circ$ ,  $90^\circ$ ,  $270^\circ$ ) to generate a single depth image. For sensors with pixels that have 2-taps (TapA and TapB), 4-frames are needed to acquire two sets of four-phase data ( $0^\circ$ ,  $180^\circ$ ,  $90^\circ$ , and  $270^\circ$  for TapA and  $180^\circ$ ,  $0^\circ$ ,  $270^\circ$ , and  $90^\circ$  for TapB)

However, the 4-frame sequence is a bit redundant and results in higher power consumption and slower depth frame rates. To address these issues, many prior works have attempted to reduce the number of frames [1-3].

## II. Design Concepts and device structure

### A. Tap mismatch

Fig. 1a shows the conventional 4-frame data readout sequence for 2-tap iToF image sensors, while Fig. 1b shows our proposed 2-frame sequence. It is usually

acquired anti-frames ( $180^\circ$  against  $0^\circ$  and  $270^\circ$  against  $90^\circ$ ) to cancel out the mismatch components between each of the taps, as illustrated in Fig. 2. This helps to reduce depth noise, particularly spatial noise (DNS). DNS is defined as the standard deviation of the depth values within a specific area after averaging the depth map of multiple frames. On the other hand, temporal depth noise (DNT) is calculated by taking the standard deviation of each pixel's depth value across multiple frames and then averaging them over the same specific area. The total depth noise can be obtained using the following formula  $\sqrt{DNS^2 + DNT^2}$ . It is worth noting that the total depth noise of the 2-frame sequence without anti-frames (Fig. 1b) will severely deteriorate if each tap has non-negligible mismatches.

The phase shift ( $\phi$ ), which is proportional to the distance, is calculated using equation (1) for the 4-frame sequence and equation (2) for the 2-frame sequence without anti-frames.

$$\phi = \text{atan} \left( \frac{(A_{90} + B_{90}) - (A_{270} + B_{270})}{(A_0 + B_0) - (A_{180} + B_{180})} \right), \quad (1)$$

$$\phi = \text{atan} \left( \frac{(A_{90} - B_{270}) + \text{mismatch}_Q}{(A_0 - B_{180}) + \text{mismatch}_I} \right), \quad (2)$$

where  $A_x$  and  $B_x$  are the sampling signals for TapA and TapB, respectively, and the subscripts 0, 90, 180, and 270 indicate each phase angles. Equation (2) suffers from the presence of mismatch components in both the numerator and denominator, which can negatively impact the quality of the resulting depth map. In this study, we aim to minimize tap mismatches to achieve high-quality 2-frame sequence for high-speed depth imaging.

### B. Device Structure and Pixel Architecture

We developed a 3D stacked BI iToF image sensor using 90 nm FEOL and 65 nm BEOL generation. This sensor has VGA resolution, with 3.0  $\mu\text{m}$  2-tap pixels. A cross-sectional SEM image of this sensor is shown in Fig. 3, which highlights the PSD structure and DTI that we incorporated into each pixel to enhance near-infrared (NIR) sensitivity [4]. Fig. 4 shows the pixel circuit, which utilizes MOS capacitors as in-pixel

memories (MEMs). This architecture enables FD sharing among adjacent  $2 \times 4$  unit pixels, making it suitable for pixel size shrinkage. FD sharing also helps to cancel the SF gain mismatch within the shared unit. After SF gain mismatch has been eliminated, other mismatches arise from TGs, MEMs, and MTRs. To address these mismatches, we adopted several technologies. Firstly, we assume that the primary source of mismatch for the TGs is the variation of carrier transfer capability between TGA and TGB. As one of the countermeasures, we designed the pixel wire routing to maximize symmetry and equalize the wiring capacitance and resistance to supply the same voltages to TGA and TGB. Secondly, to achieve high FWC, we adopted relatively large-sized MEMs, which also became the primary source of dark current and parasitic light sensitivity (PLS). We reduced the effects of dark current with process optimization. Lastly, it is crucial to have a fully carrier transfer from the MEM to FD for the MTR. Any residual carriers can cause tap mismatch; therefore, we carefully designed the MTR and optimized the space between the MTR and the MEM for smoother carrier transfer.

#### C. Dual-VG for TG mismatch mitigation

We implemented a dual-vertical gate (VG) for TG mismatch mitigation, as depicted in Fig. 5. To ensure optimal electrical potential gradient, each pixel includes a pair of VGs for TGA and TGB, with carefully adjusted VG separations based on TCAD simulation, as illustrated in Fig. 6. We assume that improving the modulation of bulk potential can lead to a reduction in tap mismatches. We were also able to effectively reduce the power consumption of high-speed modulation with lower voltage swing of 1.2 V.

#### D. Countermeasure to PLS mismatch mitigation

Most iToF sensors rely on NIR laser illumination at 940 nm, which brings PLS issues due to the low absorption coefficient of crystalline Si. To address the tap mismatch resulted from PLS, we optimized the amount of OCL offset to balance PLS between the two MEMs in each pixel, considering the diffraction of PSD structure and DTI, as well as balancing the QE and MTF (Figs. 7 and 8). Additionally, we designed the MEM's vertical potential profile of the diffusion region to be as shallow as possible and the potential barrier gradient to be steep, effectively suppressing undesired carrier injection from the photodetector to MEMs, as shown in Fig. 9.

### III. Results and Discussion

Table 1 summarizes the key characteristics of our iToF image sensor. Our sensor achieves a QE of 38% at 940 nm and a high FWC of 37 ke-. As shown in Fig. 10, we achieved demodulation contrasts ( $C_{mod}$ ) of 98%, 94%, and 88% at 20 MHz, 100 MHz, and 200

MHz, respectively. The amount of PLS is approximately -39 dB, and the mismatch between TapA and TapB is less than -50 dB over the entire image area, as shown in Fig. 11. Fig. 12 shows the single depth images taken by our ToF module under indoor (0.1 klux) and outdoor (10 klux) lighting conditions. Despite containing DNS due to tap mismatch, the average total depth noise of the selected areas was found to be 0.6% for the 4-frame sequence and 0.7% for the 2-frame sequence, showing an almost comparable depth noise quality with the 4-frame sequence under indoor conditions (Figs. 12a and b). Under high ambient illumination conditions (Figs. 12c and d), the average total depth noise of the selected areas was 0.8% for the 4-frame sequence and 0.9% for the 2-frame sequence. Notably, although the impact of tap mismatch in PLS becomes significant at the peripheral region of the image area, significant degradation in depth noise has not been confirmed (Fig. 12d). Finally, we present group depth map and point cloud data taken by our ToF module with dual frequencies of 20 MHz and 100 MHz (Fig.13). The distances of the people in the front row and the wall are approximately 1.0 m and 6 m, respectively.

### IV. Conclusion

In conclusion, we successfully developed an iToF image sensor with VGA resolution and  $3.0 \mu\text{m}$  2-tap pixels. The proposed 2-frame sequence demonstrates good depth noise performance that is almost comparable with the conventional 4-frame sequence for both indoor and outdoor conditions, owing to the carefully designed tap mismatch mitigation technologies.

#### Acknowledgement

We sincerely acknowledge all the project member of Sony Semiconductor Solutions and Sony Semiconductor Manufacturing Corporation.

#### References

- [1] M. S. Keel et al., JSSC 2020, pp. 889-897.
- [2] M. S. Keel et al., ISSCC 2021, 7.1.
- [3] J. Kang et al., VLSI Symp. 2022, C05-1.
- [4] I. Oshiyama et al., IEDM 2017, 16-4.
- [5] C. Tubert et al, ESSCJRC 2021, pp. 135-138.
- [6] Y. Ebiko et al., IEDM 2020, 33-1.
- [7] Y. Kwon et al., IEDM2020, 33-2.

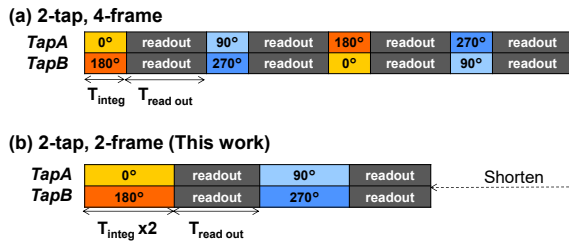


Fig. 1. The conventional 4-frame and proposed 2-frame readout sequences

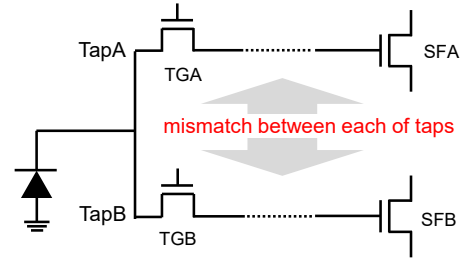


Fig. 2. Tap mismatch of 2-tap iToF pixel

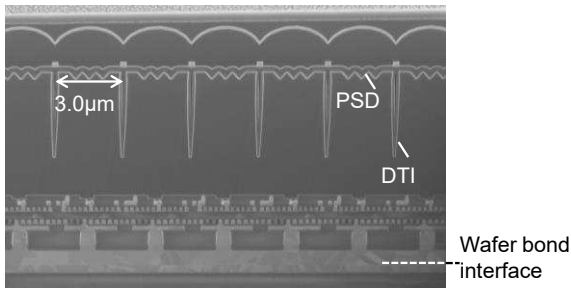


Fig. 3. The Cross Section of our device

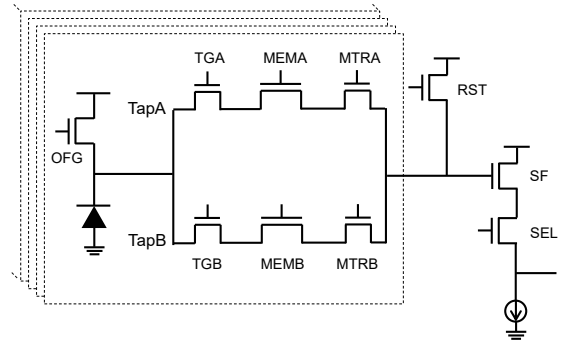


Fig. 4. Pixel architecture

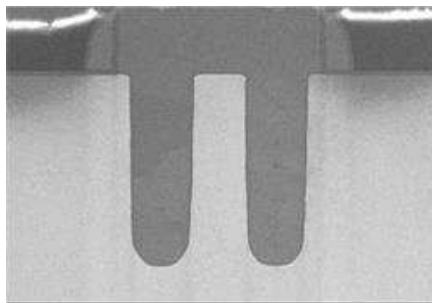


Fig. 5. Cross-section of Dual-VG

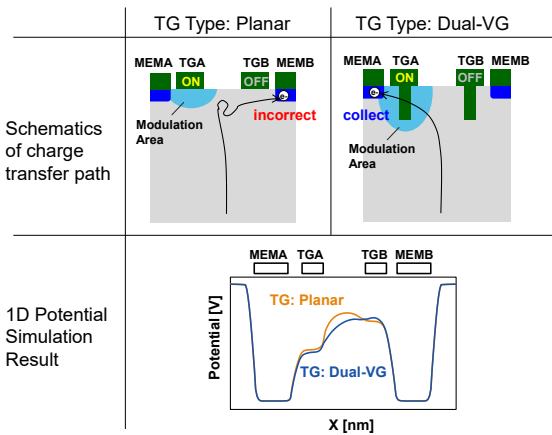


Fig. 6. Comparison between Planar TG and Dual-VG

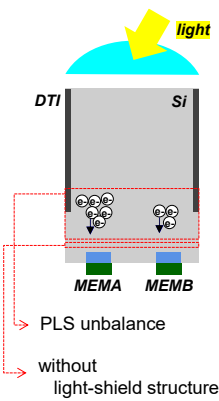


Fig. 7. The source of PLS mismatch

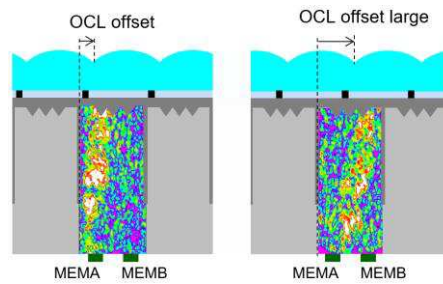


Fig. 8. OCL offset optimization and electric field distribution of 940nm light

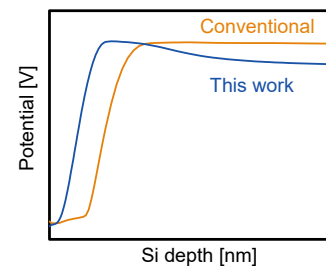


Fig. 9. Simulation results of MEM's vertical potential profile

		This work		ISSCC'21 [2]	ESSDERC'21 [5]	IEDM'20 [6]	IEDM'20 [7]
Device	Process Gen.	3D stacked BI FEOL 90nm/BEOL 65nm		3D stacked BI Top 65nm/Bottom 65nm	3D Stacked BI Top 65nm/Bottom 40nm	3D stacked BI FEOL 90nm/BEOL 65nm	BSI 65nm
	Pixel Pitch	3.0 $\mu\text{m}$		3.5 $\mu\text{m}$	4.6 $\mu\text{m}$	3.5 $\mu\text{m}$	2.8 $\mu\text{m}$
	Number of taps	2-tap		4-tap	2-tap	2-tap	4-tap
	Pixel Array	640 x 480		1280 x 960	672 x 804	1280 x 960	640 x 480
	TG Type	Dual-VG		-	-	-	-
	Charge Storage	MOS Cap.		MOS Cap.	CDTI	FD	MOS Cap.
	Frequency Modulation	10 to 200 MHz		10 to 200 MHz	up to 250 MHz	10 to 120 MHz	-
Characteristics	Demodulation Contrast	88% at 200 MHz @1.2V Swing		80% at 200 MHz @1.05V Swing	88.5% at 200 MHz @1.2V Swing	-	86% @100 MHz
	FWC	37000 e-/tap		-	-	18000 e-/tap	20000 e-/tap
	QE at 940nm	38%		38%	18.5%	32%	36%
Total depth noise	the number of acquiring frames	4-frame	2-frame	-	4-frame	-	-
	Indoor (0.1Klux)	0.6%	0.7%	-	-	-	-
	Outdoor (10Klux)	0.8%	0.9%	-	-	-	-

Table.1. Comparison of the major iToF specifications

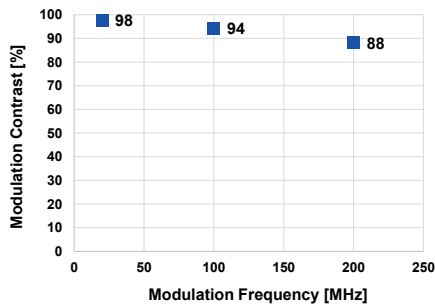


Fig. 10. Modulation Frequency dependency of Modulation Contrast

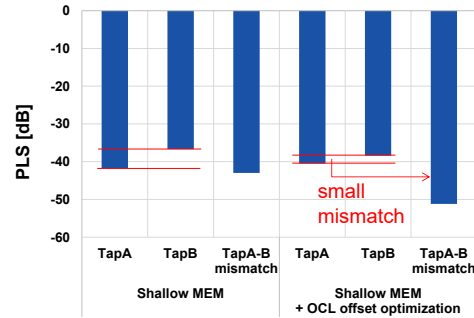


Fig. 11. PLS mismatch

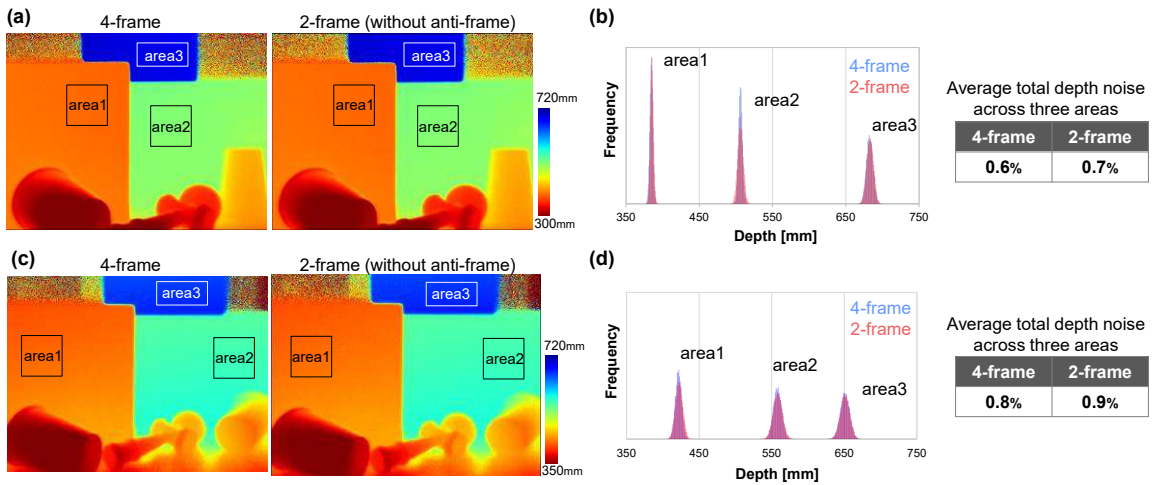


Fig. 12. Single depth images taken at indoor and outdoor (200 MHz frequency, 800  $\mu\text{s}$  total integration time)  
 (a) depth image at indoor (0.1 klux), (b) histogram of each area at indoor  
 (c) depth image at outdoor (10 klux), (d) histogram of each area at outdoor

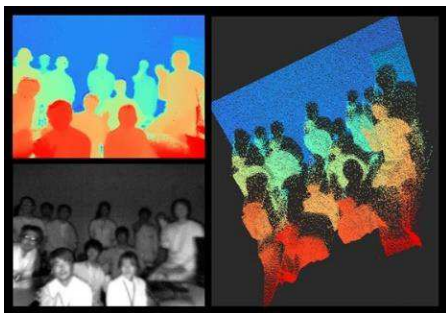


Fig. 13. Group photo taken by the sensor

# A 1.2Mp indirect-ToF sensor with on-chip ISP for low-power and self-optimization

Seung-Chul Shin, Jiheon Park, Daeyun Kim, Myoungoh Ki, Myunghan Bae, Hoyong Lee, Jonghan Ahn, Myeonggyun Kye, Bumsik Chung, Inho Song, Sunhwa Lee, Il-Pyeong Hwang, Taemin An, Jaeil An, Min-Sun Keel, Young-Gu Jin, Youngchan Kim, Youngsun Oh, Juhyun Ko, JoonSeo Yim

System LSI Division, Samsung Electronics Co., Ltd., Hwaseong, Gyeonggi-do, Korea,  
E-mail: sc1225.shin@samsung.com

**Abstract** — We propose an i-ToF to improve power consumption and motion artifact. AR/VR and Robot devices that require depth information when implementing applications have limitations when using i-ToF due to limited power and moving motion characteristics. In particular, the increase in the amount of AP operation according to the implementation of high resolution and the limitation of latency time due to multi-frame operation occur. The proposed sensor has an embedded depth processor to process depth calculation inside the sensor and enables low-power realization by using a 28nm process. In addition, motion artifact was reduced by operating at 250FPS readout speed. As a result, 30FPS depth output can be output from the sensor itself, and power consumption is reduced by 90%.

## I. INTRODUCTION

Recently, 3D information is being used as essential information for machine vision such as 3D map generation and gesture control in AR/VR and robot with RGB image. These devices require low-power operation due to its battery-based operation, and since continuous movement occurs, reduced motion and constant FPS are required. In addition, high-resolution depth information using indirect-ToF (i-ToF) is mainly considered because high-resolution depth information is required due to the characteristics of applications that use surface information. In previous studies, depth performance improvement and distance increase were proposed through demodulation contrast (DC) improvement, modulation frequency increase and tap mismatch calibration [1-4]. However, since all depth processing is software-based in the application processor (AP), there is a limit to the reduction of the overall system power. Also, depending on the performance of the AP, the FPS and depth quality is varying. In addition, since multiple sensors operate simultaneously for tracking and detection, the amount

of data transmitted to the AP and the computational burden further increase. As another issue, since the external light conditions are accumulated during the exposure time, reduced operating distance, and increased noise in outdoor environment compared to the indoor. To solve this problem, a method of converting analog binning and digital binning through external control to implement signal increase and FWC increase effects, respectively, has been proposed. [5] However, there arises a problem that a sensor for determining the external light situation is added or the operation mode needs to be changed through additional calculations in the AP. In this paper, the following three improvements are proposed sensor with the embedded depth processor unit (eDPU) built into the sensor: 1) Reduction of system power, 2) Minimization of motion artifact, 3) Reduction of influence of external conditions.

## II. BASIC OPERATION

The i-ToF is to calculate the distance through the phase delay of the modulated transmitter signal reflected back to the object. For this, phase information divided into four per cycle is required, and it is sensed by separating it into each tap inside the one pixel. This feature will reduce noise and improve accuracy. However, each tap has gain and offset errors due to the mismatch during fabrication. To compensate for this, continuous 2-frame information is generally used as tap shuffle is applied in additional frames. In addition, due to the periodicity of the modulation, the maximum operating distance is limited by frequency (fm) as a folding error. To solve this problem, the actual distance is determined through greatest common divisor (GCD) operation by cross-operating the two or more fm. As a result, when calculating i-ToF based processing, 2 to 4 multi frame based calculations are used. As a result, latency increases and the amount of required computations increases.

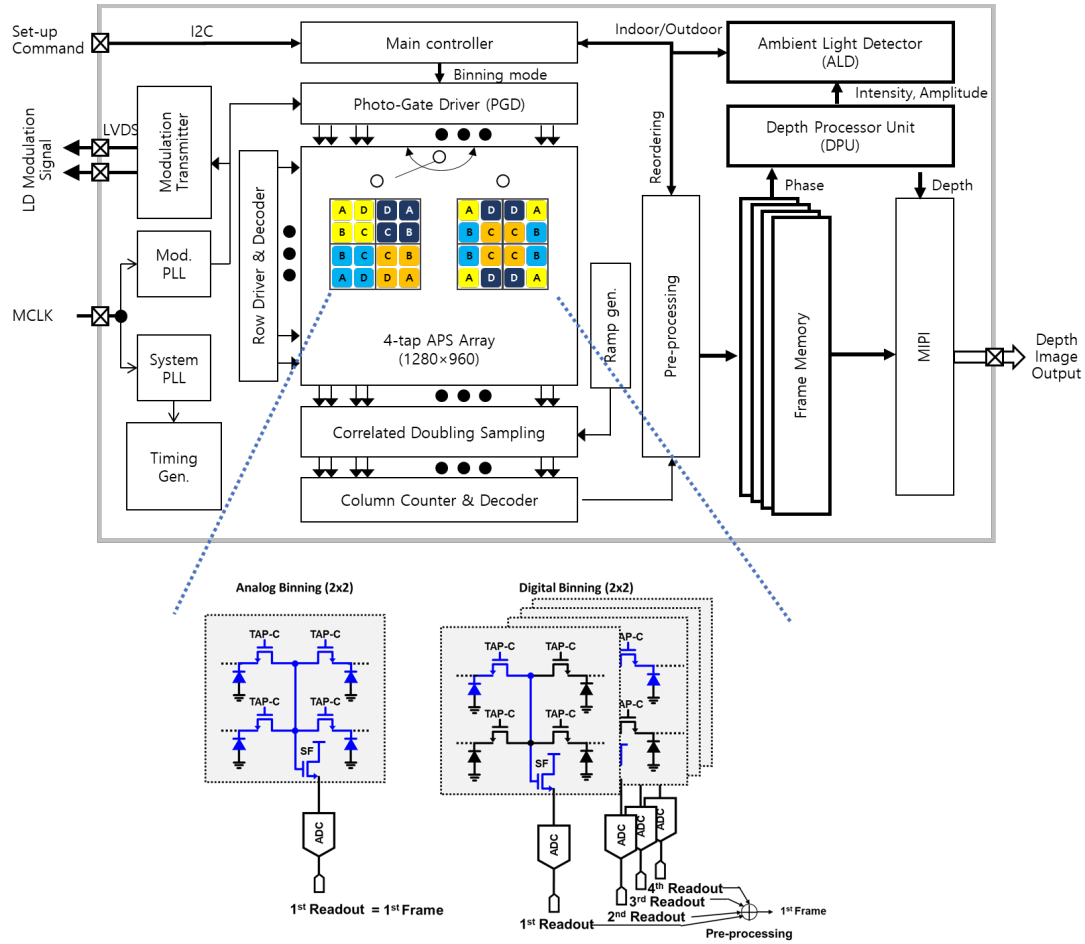


Fig. 1: Block diagram with depth processor and binning mode operation. Pixel structure is referenced [3].

### III. PROPOSED ARCHITECTURE

Fig. 1 shows a block diagram of the proposed sensor with depth processor. A unit pixel is composed of  $2 \times 2$  sub-pixels, each of which is composed of 4-tap (A, B, C, D). The phase information is stored during the exposure integrate time (EIT) section under the control of the photo gate driver signal. After EIT, analog signal is digitized using correlated double sampling (CDS) and counter. These data are subjected to phase reordering through a pre-processing block, and is stored in a frame memory consisting of up to 4-pages of raw data for shuffle and multi frequency. This signal is converted into depth information through the depth processing unit built into the sensor. The eDPU performs phase delay calculation, non-ideality and lens calibration, spatial filtering, and temporal filtering to calculate depth and enhancement. In general, the hard-wired logic has the disadvantage that only pre-defined operations are possible. However, the proposed eDPU implements simple mathematical operations (eg, phase delay sensing) with hard-wired logic, and the parts that

require function change according to the application, such as filtering, are programmable configured. The ambient light detector (ALD) calculates the ambient light environment using the intensity and amplitude calculated from the eDPU and determines the binning mode by sensor itself. This information delivers to the pre-processing block for data reordering. As shown in the bottom of Fig. 1, in case of analog binning since the summation signal of  $2 \times 2$  sub-pixels is read at once to increase signal. In case of digital binning,  $1280 \times 960$  signals are captured from all pixels to increase full well capacity (FWC) and converted into  $640 \times 480$  through average in the pre-processing block.

The timing diagram is shown in Fig.2. The proposed eDPU supports various system scenarios and minimizes motion artifacts by optimizing the raw frame rate based on shuffle, multi frequency, and ISP calculations based on depth frame. Fig.2.(a) shows the shuffle operation under single frequency modulation (s-fm). It calculates depth using two-frame and has the strength of minimizing motion.

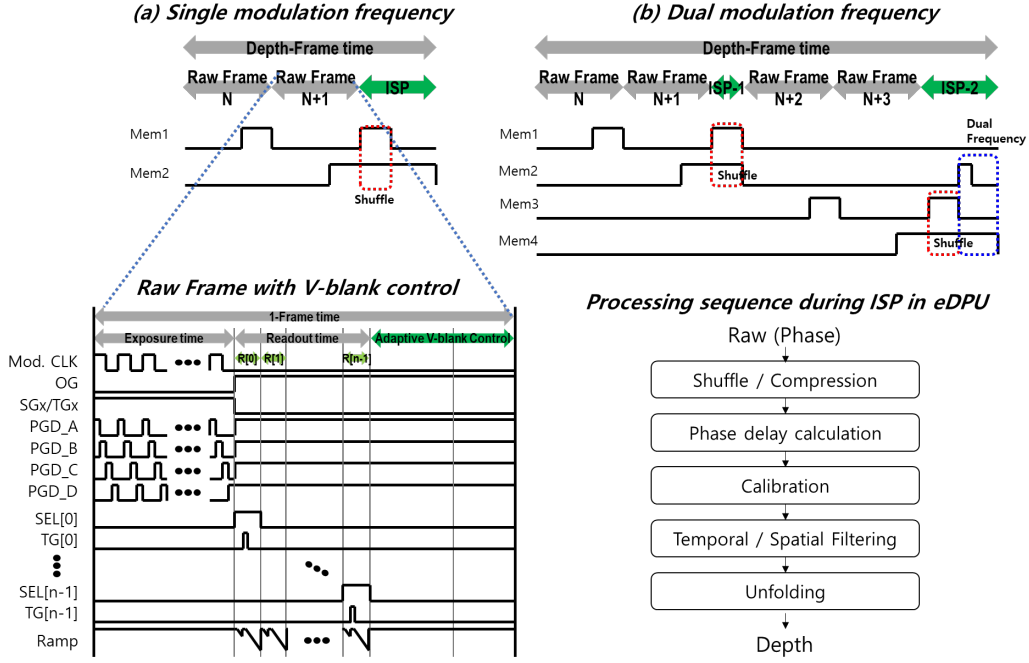


Fig. 2: Timing diagram of frame: depth frame with single and dual modulation frequency, raw frame about demodulation, and eDPU processing sequence

However, since the operating distance is determined only by the modulation frequency, it is suitable for short-distance operation. For memory size optimization, non-shuffle data ( $n$ -frame) is directly stored in frame memory, and shuffle frame data ( $n+1$ -frame) uses line memory to calculate shuffle, then compressed data stored in frame memory. Next, the depth information is output through phase delay and filtering operations. Fig.2.(b) shows the operation case of dual frequency modulation (d-fm). Although motion lagging time increases because it operates based on four-frame information, it has advantages in securing high-performance depth through distance increase and high-frequency modulation. The data ( $n$ ,  $n+1$  frame) corresponding to modulation 1st frequency ( $f_1$ ) is shuffled and calculated phase delay then stored in memory, and after reading out ( $n+2$ ,  $n+3$  frame) data of 2nd frequency ( $f_2$ ), shuffle, calibration,

filtering and unfolding operation is performed. This sequence is possible to support motion reduction and distance increase according to the application, respectively.

#### IV. IMPLEMENTATION AND RESULTS

The proposed sensor is implemented with a 2-stack structure, the top chip is a pixel array with 65nm back-side illumination (BSI), and the bottom chip has been

fabricated with a 28nm logic process for chip-size optimization due to the implementation of eDPU and frame memory. To minimize motion, raw data operates at 250 fps, and Fig. 3 shows the experimental results with rotating at 100 RPM to check the effect of motion artifact. Motion lagging time is defined from 1st raw frame start to EIT of last frame. During dual frequency operation, a lagging time of 25.39ms occurs. Fig. 3.(a) shows the experimental results of operation with dual frequency ( $f_1=100\text{MHz}$ ,  $f_2=30\text{MHz}$ ) in 60fps raw frame condition. The depth information of the moving area is distorted due to the 50.38ms lagging time, and it can be confirmed that some areas are masked due to low SNR degradation. Fig. 3.(b) shows the case of applying shuffle at  $f_m=50\text{MHz}$  with 8.73ms lagging time. It shows that the error area is reduced. Additionally, motion free performance can be secured when shuffle is excluded as shown in Fig. 3.(c), however, the noise characteristics are twice deteriorated due to tap mismatch.

Fig. 4 shows the depth and eDPU performance of the proposed sensor. For evaluation, lens is used F/1.3,  $78^\circ$  FoV, and transmitter was used with 940nm 3.4W peak power based vertical-cavity surface-emitting laser (VCSEL). The maximum EIT was set to 0.4ms, and saturation was prevented by controlling the exposure time in the main controller. For performance comparison with SW-ISP, an experiment is used with an

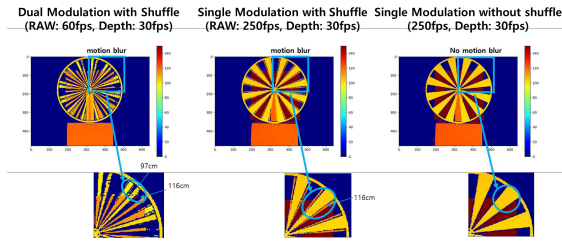


Fig. 3: Comparison of motion artifacts according to modulation frequency and shuffle application

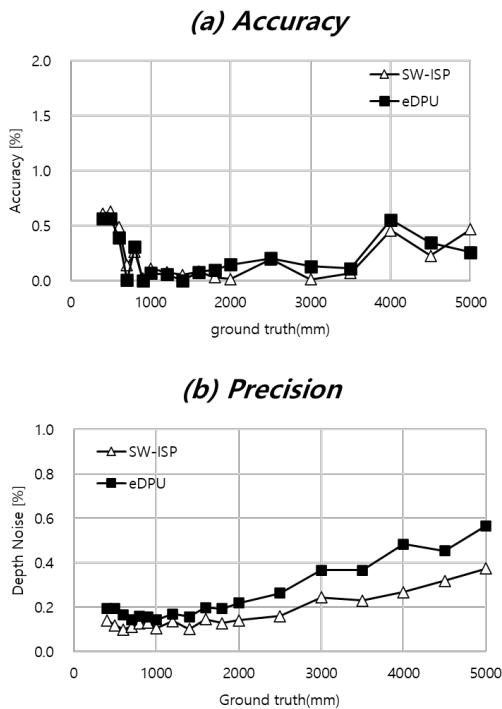


Fig. 4: Depth performance comparison of depth processing performance of AP-based SW and eDPU: Accuracy, Noise, Latency, and Power

AP Processor (HDK8350, 64bit float) by implementing the same function of 640x480 resolution. As a result of comparison, accuracy was less than 0.6% within the working range (0.4 ~ 5m). Depth noise was achieved to be less than 0.57%, but it was degraded by 0.2% compared to SW-ISP. The reason is that the quantization noise increases during the compression process for optimizing the frame memory capacity. As a result of comparison of latency time, it was to 13.59ms, and the amount of output data was also reduced to 6.25% ~ 12.5%. Processing power consumption is only 0.12W, which is reduced by 90% than SW-ISP as shown in Fig. 5. Finally, the sensor + depth operation power was reduced from 1.4W to 0.29W. Fig. 5 shows the depth performance according to the ambient condition.

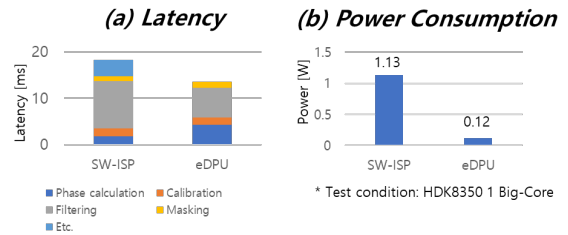


Fig. 5: Latency and power consumption comparison of depth processing performance of AP-based SW and eDPU

## V. CONCLUSION

The proposed i-ToF system enables low-power design of miniaturized devices such as AR/VR with the embedded ISP. In addition, the same performance and high-speed FPS compared to SW were confirmed. Through this, it is possible to use depth data more easily and efficiently when implementing applications and systems.

## REFERENCES

- [1] Bamji, Cyrus S., et al. "1Mpixel 65nm BSI 320MHz demodulated TOF Image sensor with 3 $\mu$ m global shutter pixels and analog binning." 2018 IEEE International Solid-State Circuits Conference-(ISSCC). IEEE, 2018.
- [2] Keel, Min-Sun, et al. "A VGA Indirect Time-of-Flight CMOS Image Sensor With 4-Tap 7- $\mu$ m Global-Shutter Pixel and Fixed-Pattern Phase Noise Self-Compensation." IEEE Journal of Solid-State Circuits 55.4 (2019): 889-897.
- [3] Keel, Min-Sun, et al. "7.1 A 4-tap 3.5  $\mu$ m 1.2 Mpixel Indirect Time-of-Flight CMOS Image Sensor with Peak Current Mitigation and Multi-User Interference Cancellation." 2021 IEEE International Solid-State Circuits Conference (ISSCC). Vol. 64. IEEE, 2021.
- [4] Kang, Jubin, et al. "A 640 $\times$  480 Indirect Time-of-Flight Image Sensor with Tetra Pixel Architecture for Tap Mismatch Calibration and Motion Artifact Suppression." 2022 IEEE Symposium on VLSI Technology and Circuits (VLSI Technology and Circuits). IEEE, 2022.
- [5] Shin, Seung-Chul, et al. "Indirect-ToF system optimization for sensing range enhancement with patterned light source and adaptive binning", Proc. IISW, 2021, [online] Available: <https://imagesensors.org/Past%20Workshops/2021%20Workshop/2021%20Papers/P12.pdf>