A 134 × 132 4-Tap CMOS Indirect Time-of-Flight Range Imager Using In-Pixel Memory Array With 10 Kfps High-Speed Mode and High Precision Mode

Chia-Chi Kuo, *Student Member, IEEE*, and Rihito Kuroda, *Member, IEEE*

***Abstract*— This article presents a prototype 4-tap indirect time-of-flight (iToF) CMOS imager with high-speed (HS) range imaging capacity. The sensor features an HS charge modulator, in-pixel memory array, and sub-frame ToF operation, enabling up to 10 Kfps range imaging with eight recording frames and <1.82% depth noise for 0.3–1.4 m range under HS mode. This sensor also operates in high-precision (HP) mode, achieving <1.77% depth noise for the 0.4–5.4 m range at 90 frames/s by averaging the subframe signals. With a pixel size of 22.4^{H} × 16^{V} μm and 134^{H} × 132^{V} 4-tap pixel array, the sensor successfully demonstrated precise depth imaging under HP mode and clear 3-D imaging for rapid motion objects under HS mode. The potential application of the sensor and future improvements are also discussed.**

***Index Terms*— 3-D Imaging, CMOS image sensor (CIS), depth sensor, high-speed (HS) imaging, indirect time-of-flight (iToF).**

I. Introduction

THE growing demand for 3-D sensing in applications such as modeling, biometrics, gesture control, and virtual reality (VR)/augmented reality (AR) has led to the development of 3-D imaging technologies with higher spatial resolution and precision. In recent years, new possibilities have emerged in the field of computer vision, including robotics, industrial automation, autonomous vehicles, and medical assistance systems. To enhance the reliability of scene analysis and decision-making for tasks such as object recognition, tracking, and navigation, depth images with higher temporal resolution are required. Moreover, achieving a performance level over 1 Kfps would enable more opportunities in advanced machine vision applications [1]. Therefore, there is a desire to develop high-speed (HS) range imagers that greatly exceed standard video rates (30/60 frames/s) to provide crucial spatial information.

A structured light algorithm as well as a direct time-of-flight (dToF) SPAD sensor has been reported for HS 3-D imaging, achieving up to 1 Kfps [1], [2]. However, the need for GPU acceleration in complex post-processing and HS readout circuits limits the feasibility of higher frame rates and image resolutions. Besides, larger module sizes and higher power consumption reduce the portability of the sensor system.

In contrast, the indirect time-of-flight (iToF) based range imagers offer the advantages, such as scalable detection range, low power consumption, cost-effectiveness, and compactness of the system [3], [4], [5], [6], [7], [8]. Nevertheless, no iToF approaches have been reported for HS 3-D imaging.

In an iToF system, it has been demonstrated that the signal-to-noise ratio (SNR), modulation frequency, and demodulation contrast (DC) are inversely proportional to the depth noise [9]. Therefore, designs with higher full-well capacity (FWC) are often preferred to achieve higher system SNR and ambient light tolerance [10], [11]. In addition, a smaller pixel pitch is also desired to improve DC when increasing the modulation frequency. However, these limit the increase in frame speed due to the longer exposure period and multi-frequency synthesis demands [12]. Consequently, current iToF imaging system designs need to compromise between these trade-offs, resulting in a confined field of applications.

Recently, a 4-tap iToF range imager enables low noise imaging at 90 frames/s and HS imaging at 2 K to 10 Kfps using an in-pixel memory array and sub-frame ToF operation was reported [13]. This sensor exhibits the potential for achieving fine depth image quality while also having the capacity to provide high temporal resolution 3-D images.

In this article, we discuss in detail the design concepts, theoretical calculations, and measured characteristics of the developed CIS for range imaging. In addition, we demonstrate the potential application of combining the high precision and HS depth imaging.

Section II describes key technologies, design concepts, and system structure of this work. Section III shows measurement results, demonstration images, and discussion. Finally, the conclusion is given in Section IV.

II. Sensor Implementation

A. Indirect-ToF Using 4-Tap Short Pulse Modulation

To address the challenges of HS imaging, multi-tap architectures have emerged as a suitable scheme to mitigate the motion artifact and enhance the depth precision [6], [14]. The 4-tap 4-phase iToF operation using sinusoidally modulated light is commonly adopted for continuous wave (CW) modulation due

Manuscript received 8 February 2023; revised 23 April 2023; accepted 22 May 2023. Date of publication 15 June 2023; date of current version 30 January 2024. This article was approved by Associate Editor Nick van Helleputte. *(Corresponding author: Chia-Chi Kuo.)*

Chia-Chi Kuo is with the Graduate School of Engineering, Tohoku University, Sendai 980-8579, Japan (e-mail: kuo.chiachi.s2@dc.tohoku.ac.jp).

Rihito Kuroda is with the Graduate School of Engineering, Tohoku University, Sendai 980-8579, Japan, and also with the New Industry Creation Hatchery Center, Tohoku University, Sendai 980-8579, Japan.

Color versions of one or more figures in this article are available at https://doi.org/10.1109/JSSC.2023.3281610.

Digital Object Identifier 10.1109/JSSC.2023.3281610

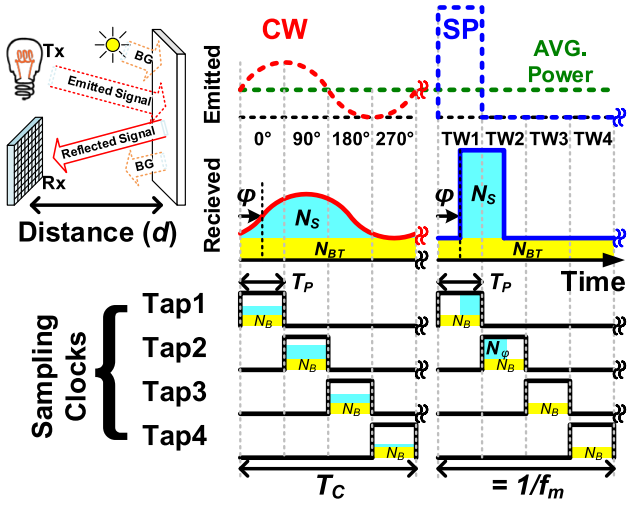


Fig. 1. Timing diagram of 4-tap iToF operation.

to its robustness and stability toward a higher modulation frequency. Furthermore, multi-frequency synthesis can be applied to extend the unambiguous range [5], [7].

In contrast, the short pulse (SP) modulation prepared by 4-tap 1-drain iToF pixel has been reported to have better ambient light tolerance due to the concentrated light energy from the emitter [15], [16]. In addition, the range-shifted technique can be adopted to increase the number of time-windows (TWs) for long-range detection [11], [17].

Fig. 1 illustrates the timing diagram of a 4-tap iToF operation using CW and SP modulation under the same emitter power condition and modulation frequency (f_m). The modulation cycle (T_C) is equally divided by four sampling clocks (T_P). The reflected signals, including the modulated light and background light, are demodulated into floating diffusions (FDs) by Tap1, Tap2, Tap3, and Tap4, and are denoted by Q_1, Q_2, Q_3, and Q_4, respectively. The distance d_{CW} and d_{SP}, with background light cancellation, can be calculated using the following equations, assuming the ideal sinusoidal and square pulse lights are given:

$$d_{\mathrm{CW}} = \frac{c}{2}\cdot\frac{1}{2\pi f_m}\cdot \mathrm{atan}\left[\frac{(Q_2-Q_4)-(Q_1-Q_3)}{(Q_2-Q_4)+(Q_1-Q_3)}\right] \quad (1)$$

$$d_{\mathrm{SP}} = \frac{c}{2}\cdot\frac{1}{4f_m}\cdot\frac{Q_2-Q_4}{Q_1+Q_2-2Q_4} \quad (2)$$

where c is the speed of light. Note that (2) is formulated for the distance range where the phase shift (φ) is within TW1.

Then, by applying the error propagation to the distance equations, the depth noise, σ_{CW} and σ_{SP}, can be expressed by the following equations [5] and [14]:

$$\sigma_{\mathrm{CW}} = \frac{c}{4\pi f_m}\cdot\frac{\sqrt{2}\cdot\sqrt{N_S+N_{\mathrm{BT}}+RN^2}}{\mathrm{DC}\cdot N_S} \quad (3)$$

$$\sigma_{\mathrm{SP}} = \frac{c}{8f_m}\cdot\frac{\sqrt{N_\varphi(1-R_S)+2\left(N_B+\mathrm{RN}^2\right)\left(1-3R_S+3R_S^2\right)}}{\mathrm{DC}\cdot N_S} \quad (4)$$

where DC is the DC and $R_S = N_\varphi/N_S$. The number of total electrons integrated in a unit pixel from the modulated

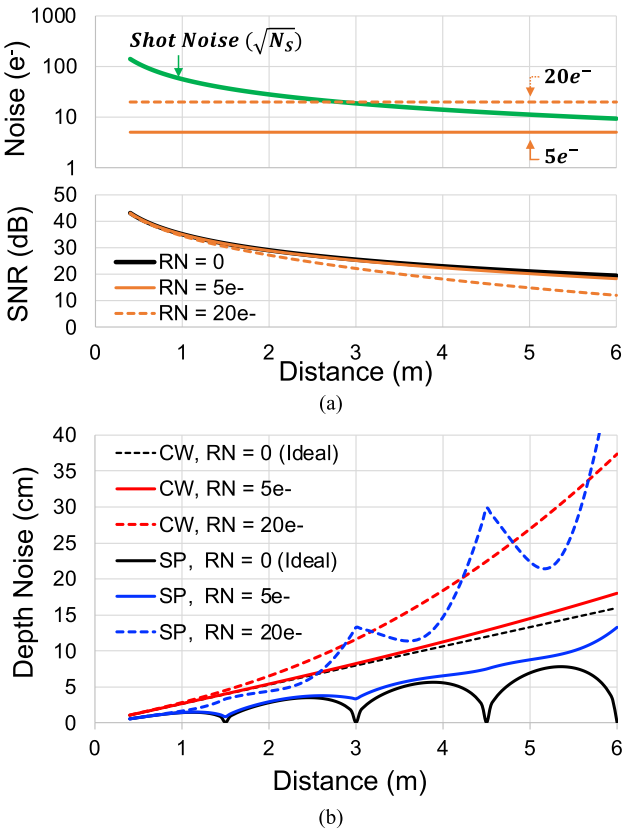


Fig. 2. (a) System noise with SNR and (b) theoretical depth noise over distance with different RN values.

light is denoted by N_S. The number of electrons generated by background light in a unit tap is denoted by N_B, and the total electrons is denoted by N_{BT}. The FD input referred readout noise (RN), which contains the pixel transistors and readout circuit, is denoted by RN. Under SP modulation, the number of signal electrons in Tap2 is proportional to the phase shift (φ), and it is denoted by N_φ.

To compare the performance of iToF operation with CW and SP modulation, we analyzed the system SNR and theoretical depth noise at different RN levels. The detection range was set to 0.4–6 m with T_p of 10 ns and T_c of 40 ns. We assumed no ambient light ($N_{\mathrm{BT}} = 0$) and assigned N_S of 20 000 at 0.4 m. The DC of both modulation methods was set to 90%, which is an achievable value in the state-of-art iToF system. The SNR is calculated by the following equation:

$$\mathrm{SNR} = 20\log\left(N_S\Big/\sqrt{N_S+\mathrm{RN}^2}\right). \quad (5)$$

Fig. 2(a) and (b) shows that SP modulation can provide better depth noise, particularity when the SNR is dominated by signal shot noise. However, it is important to note that in a scenario with higher RN, the SNR degrades more rapidly, leading to a diminished advantage when using SP modulation.

Therefore, to achieve lower depth noise and higher frame rates, a desirable solution is the development of a 4-tap iToF image sensor using SP modulation with enhanced SNR. In this work, to minimize the image processing effort and maximize the frame speed, a single frequency SP iToF system is employed without drain gate and range-shift technique.

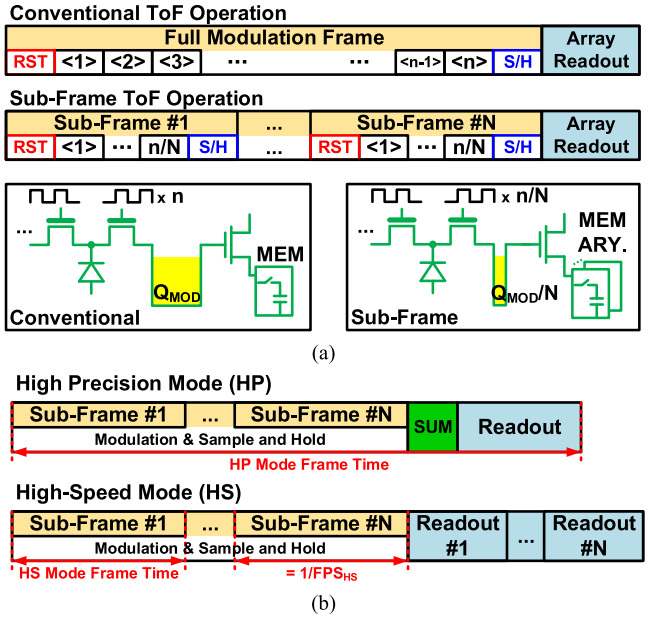


Fig. 3. Operation principle of (a) sub-frame ToF modulation and (b) HP and HS mode readout.

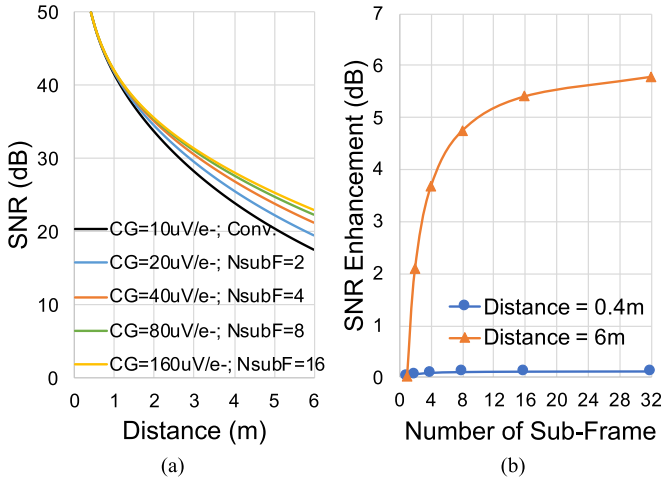


Fig. 4. Calculated characteristic with different numbers of subframes. (a) SNR over distance. (b) SNR enhancement.

B. Sub-Frame ToF Operation

To improve the system SNR of iToF imagers, increasing the FWC and exposure time are common strategies, but they come at the expense of higher input referred noise due to a lower conversion gain (CG) caused by higher FD capacitance. In contrast, pixel binning techniques, such as charge domain, analog domain, and digital domain binning can also be useful in reducing the RN but at the cost of reducing spatial resolution [18]. However, these approaches require longer frame-to-frame latency while increasing the pixel counts, making them unsuitable for HS imaging.

In this work, we introduce the sub-frame ToF operation with an in-pixel memory array to enhance the SNR while keeping the capacity of high temporal resolution imaging. As shown in Fig. 3(a), the full modulation period is divided into subframes, and the signals are sampled into individual memory cells. Subsequently, by applying different readout timings, as shown in Fig. 3(b), two types of imaging modes can be performed.

1) HP Mode: In HP mode, the modulated subframes are combined using charge-domain binning by mixing the signal charges in the memories. Owing to the averaging effect, the noise can be reduced including shot noise, flicker/thermal noise of pixel transistors, and kTC noise of the in-pixel memory. The theoretical SNR is expressed by the following equation:

$$\mathrm{SNR} = 20\log\frac{N_S}{\sqrt{\left(N_S + N_{\mathrm{PIX}}^2 + N_{\mathrm{MEM}}^2\right)/N_{\mathrm{Sub}F} + N_{\mathrm{CKT}s}^2}} \left(N_{\mathrm{PIX}} = \frac{\sigma_{\mathrm{PIX}}}{\mathrm{CG}};\ N_{\mathrm{MEM}} = \frac{\sigma_{\mathrm{MEM}}}{\mathrm{CG}};\ N_{\mathrm{CKT}s} = \frac{\sigma_{\mathrm{CKT}s}}{\mathrm{CG}}\right) \tag{6}$$

where $N_{\mathrm{Sub}F}$ is the number of subframes, CG is the CG, and N_{PIX}, N_{MEM}, and $N_{\mathrm{CKT}s}$ represent the number of noise electrons from pixel, in-pixel memory, and other readout circuits, respectively. These can be obtained by dividing the FD referred noise voltage σ_{PIX}, σ_{MEM}, and $\sigma_{\mathrm{CKT}s}$ by CG.

The calculated SNR characteristic is shown in Fig. 4(a), where a reference is established using the conventional operation with a CG of 10 $\mu\mathrm{V/e}^-$ for comparison with the subframe operation. A constant $\mathrm{CG}/N_{\mathrm{Sub}F}(\mu\mathrm{V/e}^-)$ of 10 and $N_S{\cdot}N_{\mathrm{Sub}F}(\mathrm{e}^-)$ of 100 000 are used to have an equal amount of signal electrons under a certain exposure condition. Here, σ_{PIX} of 300 μV, σ_{MEM} of 370 μV, and $\sigma_{\mathrm{CKT}s}$ of 200 μV are assumed.

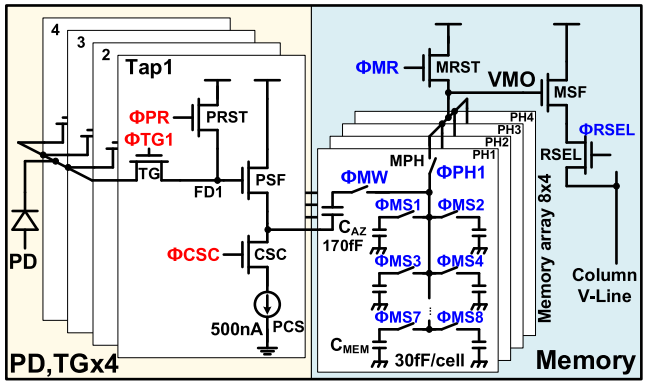
Fig. 4(b) depicts the SNR enhancement achieved through HP mode. Minor improvement was observed in shorter distances due to shot noise dominance. As the distance increases, a higher number of subframes with higher CG results in better SNR due to the reduction of equivalent noise electrons. The optimal number of subframes is eight, considering SNR enhancement and hardware complexity.

2) HS Mode: HS mode readout the memory cells individually, enabling the acquisition of high temporal resolution images. In this burst imaging mode, the frame rate is defined as one divided by the subframe period, and the available record length is determined by the number of subframes. The SNR is expressed by the following equation:

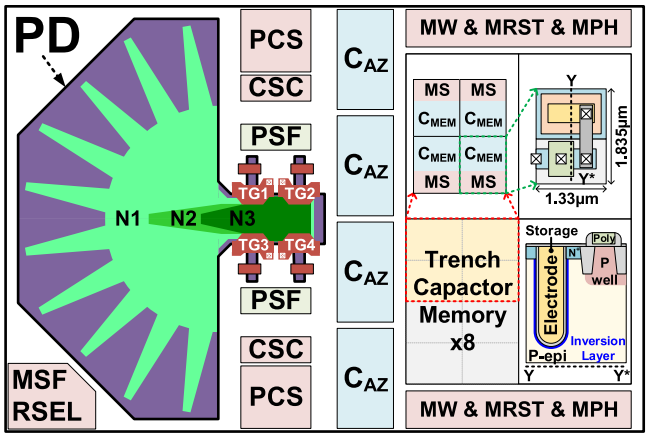
$$\mathrm{SNR} = 20\log\frac{N_S}{\sqrt{N_S + N_{\mathrm{PIX}}^2 + N_{\mathrm{MEM}}^2 + N_{\mathrm{CKT}s}^2}}. \tag{7}$$

C. Circuit Architecture and Operation

In this prototype sensor, a 4-tap iToF pixel was developed with eight subframes. To reduce the sampling period and minimize the exposure deadtime between subframes, an in-pixel memory array was employed. Fig. 5(a) and (b) shows the diagrams of the pixel, which consists of a 4-tap modulator with HS charge collection photodiode (PD), demodulation gates (TG1-TG4), four sets of buried channel source follower (PSF), current source (PCS) with cascode switch (CSC), auto-zeroing capacitor (C_{AZ}), and 4 × 8 1-T 1-C analog memory array with control devices [memory write (MW), memory reset (MRST),



(a)



(b)

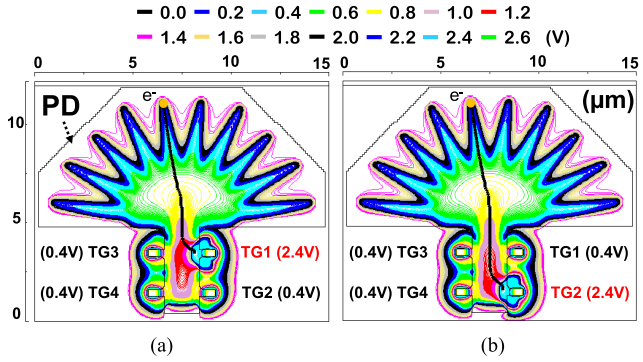
Fig. 5. Developed 4-tap iToF pixel with in-pixel memory. (a) Circuit diagram. (b) Layout diagram.

memory phases (MPHs)], which share a column readout buffer [memory source follower (MSF), row select (RSEL)]. The analog memory capacitor (C_{MEM}) and C_{AZ} utilize high-density Si trench capacitors [19], [20], providing capacitance of 30 and 170 fF, respectively.

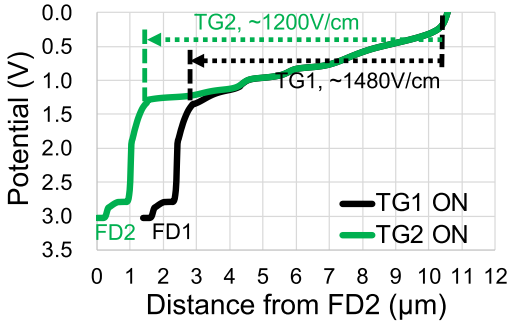
To reduce the subframe period and ensure good DC, the PD was designed with the following concepts: 1) increasing the size of PD to increase the number of generated charges; 2) enlarging the n-layer region to enhance the charge collection efficiency; 3) creating a dopant concentration ladder to generate an electric field toward the center of modulator [21]; and 4) implementing a compact 4-tap structure to improve the electron sorting efficiency by utilizing fringe electric field [22].

In this work, we also utilize a spiky triangular-shaped layout [23], [24] with three levels of n-layer (N1-N3) to create a linear potential gradient along the charge transfer path. This ensures that the generated electrons can be driven efficiently regardless of their initial position.

Simulated potential diagrams of electron transport to FD1 and FD2 are shown in Fig. 6(a) and (b), respectively. In the simulation, electrons were initially placed at a depth of 1 μm from the far end of the PD. Fig. 6(c) provides a comparison of their transfer paths. To reduce power consumption, the TG-ON voltage of 2.4 V and TG-OFF voltage of 0.4 V were used while maintaining good electron sorting efficiency. The HS charge modulator can collect charges within 0.8 ns owing to an electric field of over 1200 V/cm across the PD to the farthest FD, FD2, and FD4.

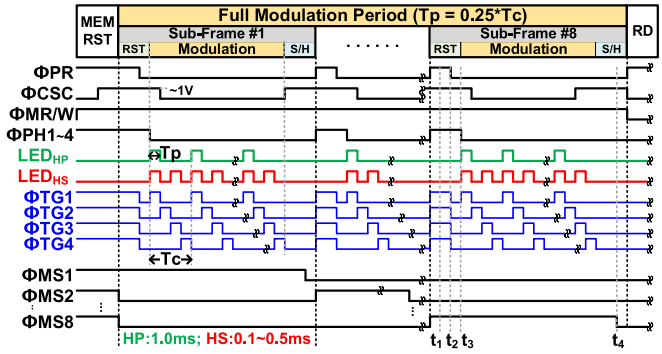


(a) (b)

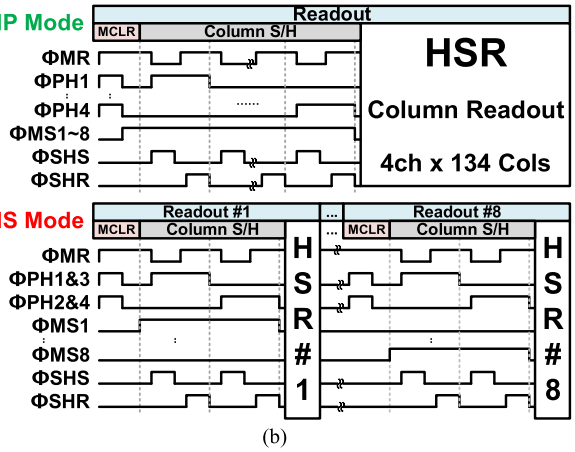


(c)

Fig. 6. Simulated potential diagrams with electron transfer path when (a) TG1 turn on, (b) TG2 turn on, and (c) 1-D plot comparison.



(a)



(b)

Fig. 7. Timing diagrams of (a) sub-frame ToF operation and (b) column readout operation of HP and HS readout.

Fig. 7(a) shows the detailed timing diagram of the sub-frame operation, constructed using eight subframes with SP modulation. At the beginning of each subframe ($t1$), the FDs,

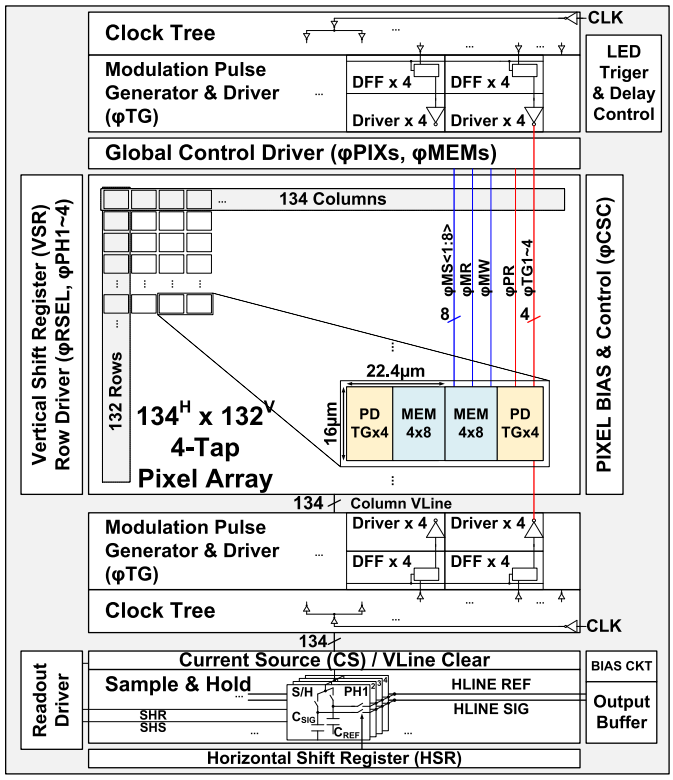


Fig. 8. Functional block diagram of the prototype sensor.

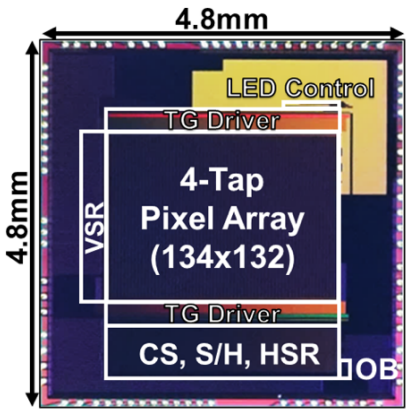


Fig. 9. Chip micrograph.

TABLE I
CHARACTERISTIC SUMMARY OF MODES

	HCG	LCG	HP	HS
CG ($\mu V/e^-$)	84.8	10.6	84.8	84.8
FWC (e^-)	12k	96k	12k	12k
Expo. (a.u.)	x1	x8	x8	x0.1 - 0.5
RN (e^-)	10.4	83.2	4.8	7.9

C_{MEM}, and C_{AZ} are reset. Auto-zeroing is then performed by ΦPR and ΦPH in sequence ($t2$, $t3$) to eliminate the thermal noise and offset from FD reset. After the modulation is completed, the 4-tap signals are simultaneously sampled into the corresponding memories by ΦMS (t4). The auto-zeroing operation reduces the required number of voltage samplings by half, resulting in a smaller memory array. It is worth noting that during the modulation, PCS is turned off by ΦCSC to reduce the power consumption.

To optimize the HP and HS modes, different modulation types and readout timing are applied, as shown in Fig. 7(a) and (b), respectively.

1) HP Mode: In HP mode, a conventional 4-tap iToF modulation (LED_{HP}) is applied to extend the unambiguous range. In addition, to achieve higher SNR for lower depth noise, the signal charges in the 8-memory of each tap are summed up before the column sample/hold (S/H) operation.

2) HS Mode: In HS mode, a pseudo-2-tap iToF modulation (LED_{HS}) is employed to increase the system SNR, despite having a shorter detection range. This is accomplished by doubling the light pulse frequency, which is demodulated alternatively by TG1/TG2 and TG3/TG4. Consequently, the selected memories in PH1 and PH3 and PH2 and PH4 are mixed before the column S/H, completing the pseudo-2-tap operation. Finally, the memory signal of each subframe is readout in sequence to obtain burst images.

For both modes, the delta-double sampling (DDS) readout [25] is performed at voltage of memory output (VMO) to reduce both fixed pattern noise and low frequency noise due to MSF readout.

Fig. 8 illustrates the functional block diagram of the proposed iToF imager, which consists of a $134^H \times 132^V$ pixel array with a pixel size of $22.4^H \times 16^V$ μm. The demodulation clocks (TGs), generated by shift registers, are driven by two sets of drivers from both top and bottom sides of the pixel array to minimize TG pulse distortion under high-frequency modulation. The LED trigger and delay control circuits are implemented to fine-tune the edge of modulation light pulse. During the rolling readout, the signals in the pixel memory arrays (MEM) are sampled by the column SH (CSH) and selected by horizontal shift register (HSR). Finally, the analog output signals from the output buffer (OB) are quantized by an off-chip 14-bits ADC.

III. MEASUREMENT RESULTS

Fig. 9 shows the micrograph of the developed prototype chip. It was fabricated using a 0.18-μm 1-poly-Si 5-metal CIS process technology with 8-μm-thick P-epi on N-sub wafer. The power supply voltage is 3.3/2.8/2.4 V for analog/digital/TG circuits, respectively. The die size is 4.8 $mm^H \times 4.8$ mm^V.

In this prototype sensor, a CG of 84.8$\mu V/e^-$, and an FWC around 12ke^- were confirmed at FD1. To evaluate the proposed HP and HS mode, conventional operation with high CG and low CG, denoted by HCG and LCG, respectively, are used for comparison. The basic characteristics are listed in Table I. Note that the results of LCG were calculated by a factor of 8.

Fig. 10 compares the SNR characteristic of HCG, LCG, and HP over different illuminance levels, indicating the maximum achievable SNR at different distances (d) in an iToF system.

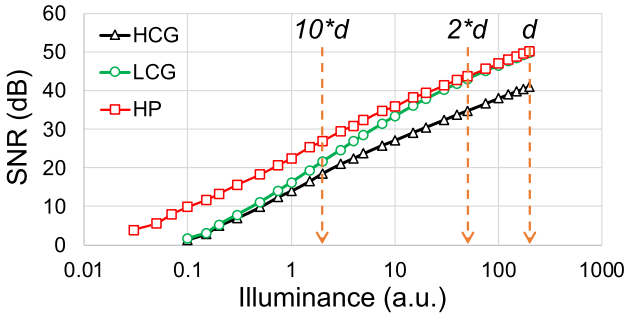


Fig. 10. SNR characteristic comparison.

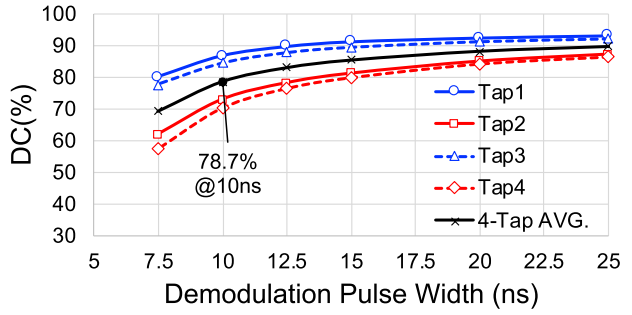


Fig. 11. DC measurement results of each tap with different demodulation pulse widths (T_P).

The LCG exhibits an advantage over HCG in higher illuminance conditions, where shot noise dominates the SNR, due to its higher FWC. In contrast, the HP mode provided over 8.4 dB SNR enhancement across the measured illuminance range, contributing to better depth noise performance in iToF range imaging. In HP mode, the total FD input-referred noise was 407 μV, consisting of the noise from pixel circuits (254 μV) and readout circuits (318 μV).

The DC characteristic of the proposed 4-tap iToF sensor with SP modulation has been evaluated and plotted in Fig. 11 with different demodulation pulse widths (T_P), and was calculated by the following equation [16]:

$$\mathrm{DC}_n = \max\left(\frac{2S_n - S_{\mathrm{SUM}}}{S_{\mathrm{SUM}}}\right), \text{ where } S_{\mathrm{SUM}} = \sum_{n=1}^{4} S_n \quad (8)$$

where the demodulated signal in each tap is denoted by S_n.

A 4-tap averaged DC of 78.7% at 10 ns demodulation pulse was obtained by the averaged data from 10 × 10 pixels in the center of the array. However, there is a DC difference of around 14% between Tap1/3 and Tap2/4 due to the asymmetrical structure of the charge modulator. The DC is strongly related to the charge collection efficiency of each Tap. Tap2/4 requires a longer electron transfer time due to a weaker electric field with longer transfer path, resulting in a lower DC. Hence, assuming that the transfer time difference remains constant, decreasing the T_P exacerbates the DC difference between Taps.

The depth performance was evaluated by analyzing the central 100 pixels over 100 consecutive frames. The system was set up with a 90% reflectivity white flat board as the target, and an 850 nm vertical-cavity surface-emitting laser (VCSEL) with a peak power of 8-W generating the modulation light pulse with a T_P of 10 ns. An F/1.4 lens with an IR bandpass filter was used, and the measurement was conducted indoors with <500 lx fluorescent light. The detailed system parameters are summarized in Table II.

TABLE II
SYSTEM PARAMETERS FOR DEPTH MEASUREMENT

	Parameter	Value
Chip	Process	0.18 µm 1P5M FSI CIS
	Array size	134 × 132
	Pixel size	22.4µm × 16µm with a fill factor of 21.6%
	Exposure time	1.0-ms / subframe @ HP mode[1] 0.5-ms / subframe @ HS mode[2]
VCSEL	Emitter type	VCSEL with a FoI of 60° × 45°
	Wavelength	850 nm with a bandwidth of 2 nm
	Optical power	8-W peak power 1.44-W avg. power @ HP mode[1] 0.50-W avg. power @ HS mode[2]
Lens	Optical filter	82.6% transmission @ 850 nm
	F#	1.4
	Focal length	25 mm

[1]At 90fps; [2]At 2 Kfps with 33.3ms refresh period;

The modulated light and TG pulses were set to 10 ns, which allowed for an unambiguous range of 1.5 m with a single TW. The HP operation provides 4-TW with the cycle time (T_C) of 40 ns, whereas the HS mode has only 1-TW due to the pseudo-2-tap operation. A time delay of ~1 ns existed between TG1 pulse and the light pulse. The slope and time offset of the measured depth were calibrated for each TW [26].

A. HP Mode

The theoretical depth noise of the proposed HP mode was compared. The equation can be expressed by the following equation:

$$\sigma_{HP} = \frac{c}{8f_m}\frac{\sqrt{\frac{N_\varphi(1-R_S)}{N_{\mathrm{Sub}F}} + 2\left(\frac{N_B}{N_{\mathrm{Sub}F}} + \mathrm{RN}_{\mathrm{HP}}^2\right)\left(1 - 3R_S + 3R_S^2\right)}}{\mathrm{DC}\cdot N_S} \quad (9)$$

where the measured RN is denoted by $\mathrm{RN_{HP}}$. Here, f_m of 25 MHz, N_B of 0, $\mathrm{RN_{HP}}$ of 4.8, $N_{\mathrm{Sub}F}$ of 8, DC of 0.86 for TW2/4, and 0.72 for TW1/3 were applied. At the distance of 0.4 m, N_S of 20 000 was used, with 12 000 electrons in Tap1 and 8000 in Tap2.

Fig. 12(a) and (b) shows the measured results of depth accuracy and depth noise, respectively. The system is capable of measuring distances ranging from 0.4 to 5.8 m at a 90 frames/s framerate with an exposure time of 1.0 ms per subframe. The measured depth nonlinearity was <1.62%, which was attributed to tap mismatches, non-ideal distortion caused by the limited bandwidth of the light pulse, and reflected stray light from the measurement system. The depth noise was measured to be <1.77% for the range within 5.4 m. Compared to the conventional HCG and LCG operation, the proposed HP mode provides a better depth noise performance across all ranges.

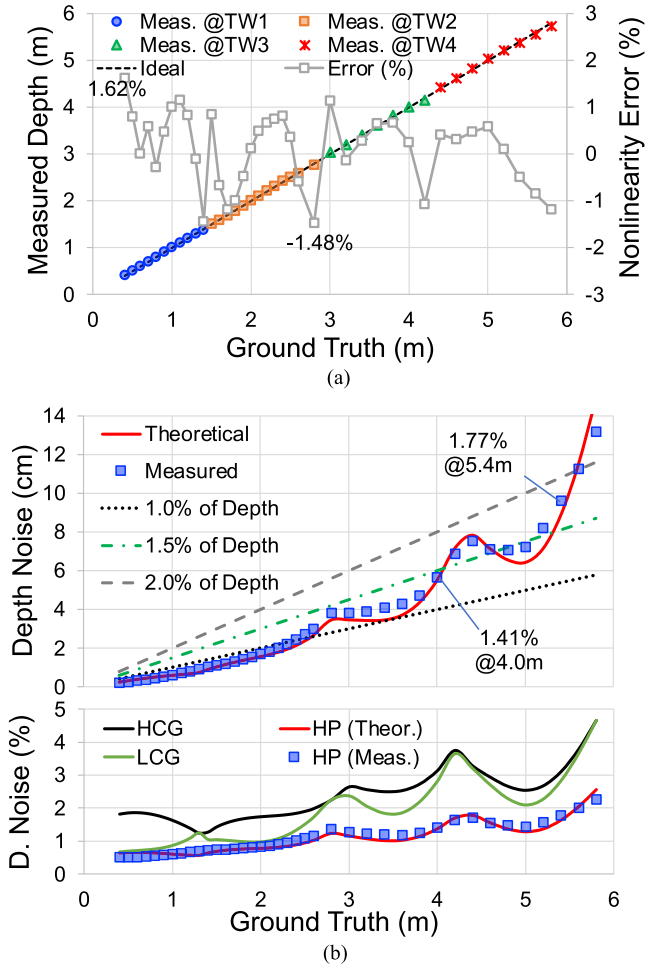


Fig. 12. Depth performances of HP mode. (a) Depth accuracy. (b) Depth noise and comparison with HCG/LCG.

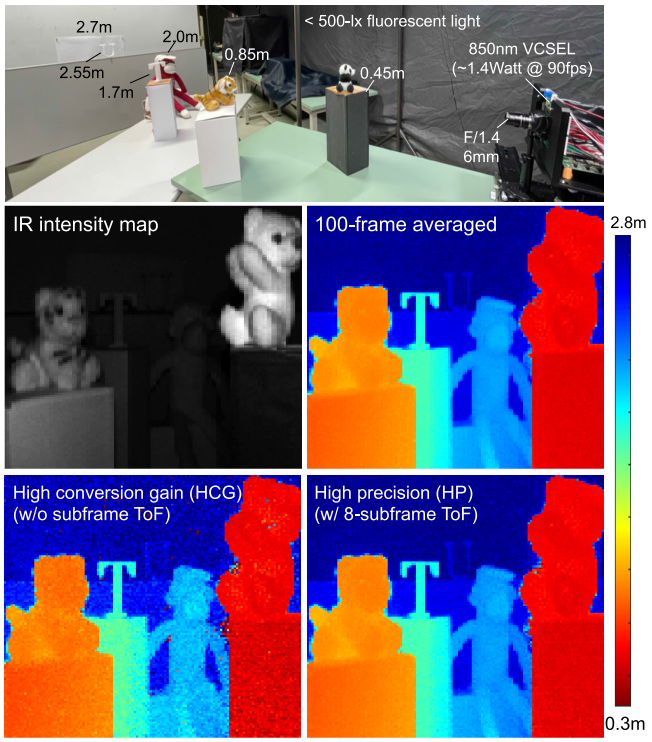


Fig. 13. Captured sample images under HCG/HP mode.

The captured sample images and environment setup are depicted in Fig. 13. The nearest object was placed at 45 cm, and an alphabet "*U*" was positioned at 2.55 m, which was 15 cm in front of the background panel. While the intensity map demonstrated a rapid decrease in reflected light as the distance increased, the proposed HP mode still delivered fine depth resolution and a distinct image of "*U*" without requiring frame averaging. The depth error observed at the edge region surrounding the objects was attributed to the "flying pixel" effect caused by spatial sampling issues [27].

B. HS Mode

The equation of HS mode, which was calculated with pseudo-2-tap operation, can be expressed by the following equation:

$$\sigma_{\mathrm{HS}} = \frac{c}{4f_m} \frac{\sqrt{\frac{N_\varphi(1-R_S)}{2} + \left(\frac{N_B}{2} + \mathrm{RN}_{\mathrm{HS}}^2\right)\left(1 - 2R_S + 2R_S^2\right)}}{\mathrm{DC} \cdot N_S} \tag{10}$$

where the measured RN is denoted by $\mathrm{RN_{HS}}$. Here, f_m of 50 MHz, N_S of 16400, N_B of 0, $\mathrm{RN_{HS}}$ of 7.9, and DC of 0.72 were applied.

Fig. 14(a) and (b) shows the depth measurement results obtained at 2 Kfps with 0.5-ms exposure time. The shorter exposure time allows for a closer measurable range of 0.3 m without signal saturation. The depth performance of each subframe was characterized separately for the range of 0.3 to 1.4 m. The 8-subframe averaged depth nonlinearity was $<$1.91%, with depth noise $<$1.82%.

Fig. 15 demonstrates depth imaging at 2 Kfps, where a spinning alphabet "*T*" at ~4500 rpm blocked the objects behind it. By using the HS mode, the captured frame could be disassembled into eight burst frames with a temporal resolution of 0.5 cms, enabling clear depth images of the background objects and observing target. Note that the burst images in HS mode are readout and refreshed every 33.3 ms. Sample images with a higher speed of 5 and 10 Kfps are to be shown in Fig. 16.

The calculated noise curves for both HP and HS modes provide a reliable estimate of the expected depth noise based on the measured system parameters. However, it should be noted that in practical iToF measurement, the depth performance is affected by several system limitations and uncertainties, such as clock distortion and jitter, limited bandwidth of the modulated light pulse for SP modulation, unstable sensor and emitter power, and heat buildup during the sensor operation. To obtain relatively stable measurement results, external voltage sources and ADC were used for this prototype sensor. Meanwhile, a heat sink was attached behind the designed emitter module.

In real practice, this development is suitable for constructing 3-D environments for human behavior analysis using HP mode, thanks to its low depth noise, suppressed

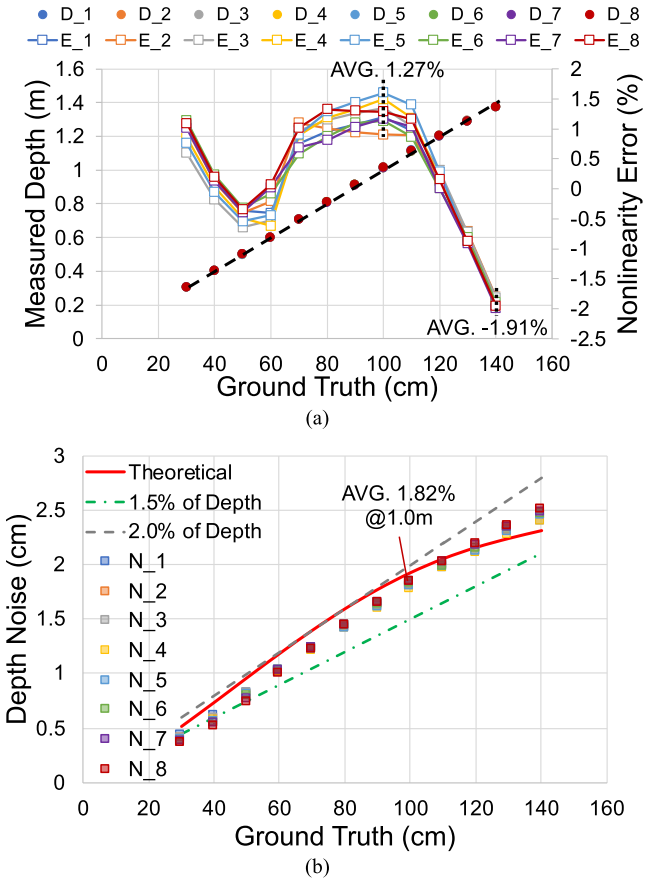


Fig. 14. Depth performances of HS mode. (a) Depth accuracy. (b) Depth noise.

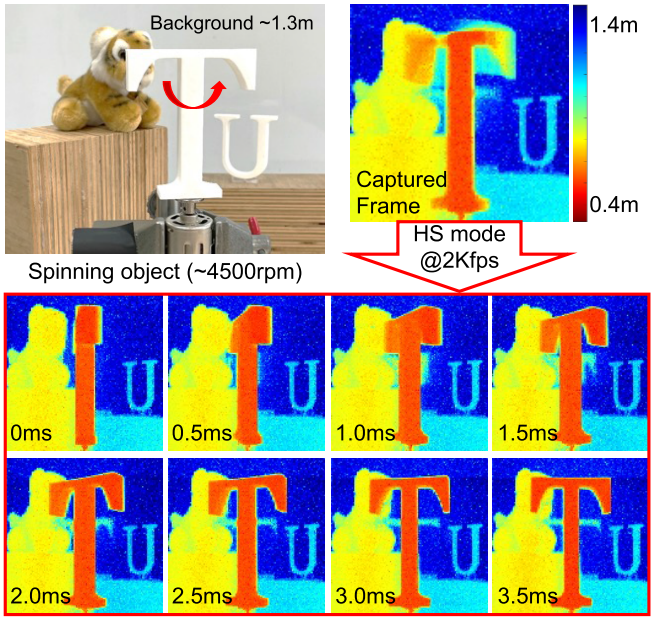


Fig. 15. Captured sample images under 2 Kfps HS mode.

motion artifact, and wide-range imaging capability. In contrast, HS mode provides high temporal resolution images that can be applied in machine vision for rapid motion recognition and analysis. Furthermore, the combination of HP and HS mode has a great potential to excel in event-driven applications, for example, a comprehensive automotive safety system.

In regular circumstances, the HP mode can continuously monitor the driver's behavior for gesture control and alert them in case of any signs of impairment, while the HS mode can be activated in case of any sudden impact or collision. In the event of severe accidents, where the deceleration forces can exceed 50G and cause the human body to come to a sudden halt within a couple milliseconds. At this moment, an HS burst imaging is needed to record the crucial spatial information.

In the experiment shown in Fig. 16, the accident was indicated by a bouncing plate using rubber bands, which was captured under 10 and 5 Kfps, with record durations of 0.8 and 1.5 ms, respectively. The pattern on the target, and its displacement and rotation, were all successfully observed.

The captured burst images can then be used to analyze the rapid motion and detect any potential injuries that might have occurred. In addition, the event-driven HS mode can quickly capture and analyze the relevant data without requiring continuous operation at such high framerate, reducing power consumption, and extending the system lifespan.

Table III compares the performance of the developed prototype sensor to state-of-the-art iToF sensors. The figure-of-merit (FoM) can be calculated using the following equation, which is another expression of [7, (3)]:

$$\text{FoM} = \frac{\text{power} \times \text{depth noise}}{\text{pixel rate}(PR) \times N_{\text{Tap}}} \left[\frac{pJ}{\text{pixel}}\right] \tag{11}$$

where the depth noise (%) is the maximum value measured in the selected range, and pixel rate (PR) is calculated using (12). In the HS mode, burst images are refreshed with a period of 33.3 ms. Hence, the equivalent PR is expressed in the following equation:

$$\text{PR} = \text{frame rate} \times \text{pixel count} \left[\frac{\text{pixel}}{s}\right] \tag{12}$$

$$\text{PR}_{\text{HS}} = \frac{N_{\text{Sub}F} \times \text{pixel count}}{\text{refresh period}} \left[\frac{\text{pixel}}{s}\right]. \tag{13}$$

In an iToF ranging system, the modulation period is adjusted based on the signal saturation level at the minimum distance. However, the power of the modulated light decreases rapidly with depth, limiting the maximum distance due to low SNR. Hence, the detection range of the measured data should be taken into consideration. For comparison, a range-FoM (R-FoM) is defined as in the following equation:

$$R - \text{FoM} = \frac{\text{FoM}}{\text{Detection range ratio}} \left[\frac{pJ}{\text{pixel}}\right] \tag{14}$$

where range ratio is ($\text{Distance}_{\text{MAX}}/\text{Distance}_{\text{MIN}}$) of the selected range, and the $\text{Distance}_{\text{MAX}}$ is chosen based on the minimum achievable R-FoM, which indicates the optimal working range with the highest efficiency for an iToF imager.

This prototype sensor achieves an R-FoM of 16 and 18.4 pJ/pixel in HP and HS mode, respectively, for a range of 0.4–5.4 and 0.3–1.4 m.

For future development, the performances can be improved from several aspects. First, sensor resolution can be significantly enhanced by introducing the backside illumination

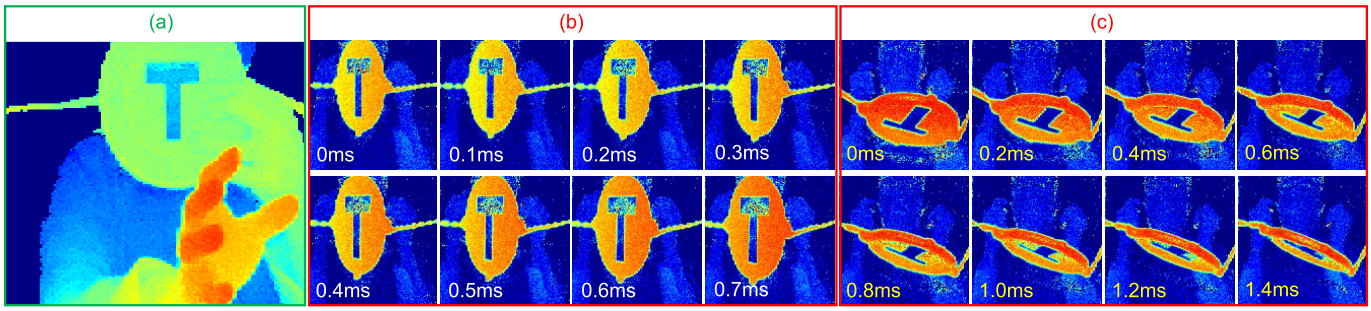


Fig. 16. Demonstration of a scenario for the automotive safety system using the proposed sensor with (a) HP mode at 90 frames/s, (b) HS mode at 10 Kfps, and (c) HS mode at 5 Kfps.

TABLE III
PERFORMANCES SUMMARY AND COMPARISON

	This Work		JSSC '21 [7]	JSSC '22 [8]	JSSC '23 [11]	JSSC '20 [5]	JSSC '20 [6]	JSSC '15 [12]
Process	0.18μm FSI		65nm BSI	110nm BSI	110nm BSI	65nm BSI	90nm BSI	130nm FSI
Pixel pitch(μm)	22.4 × 16		3.5 × 3.5 (stack)	7.2 × 7.2	5.6 × 5.6	7 × 7 (stack)	8 × 8	10 × 10
Pixel archit.	4-tap		4-tap	2-tap	4-tap	4-tap	2-tap	2-tap
Pixel array	134 × 132 [1]		1280 × 960	640 × 480 [1]	640 × 480	640 × 480	320 × 240	512 × 424
Modulation speed	25MHz (HP), 50MHz (HS) (10ns short-pulse)		10-200 MHz (sine CW)	75 MHz (pulse CW)	17ns (short-pulse)	10-150 MHz (sine CW)	10-100 MHz (pulse CW)	10-130 MHz (pulse CW)
Responsivity	0.096 A/W @ 850nm		0.288 A/W @ 940 nm	- @ 940 nm	0.243 A/W @ 940 nm	0.258 A/W @ 940 nm	- @ 855 nm	0.144 A/W @ 860 nm
Demodulation contrast	78.7% [2] @ 10ns PW		80% @ 200MHz	60% @ 75MHz	94% @ 17ns PW	86% @ 100MHz	61.2% @ 100MHz	68% @ 50MHz
Conv. Gain	85μV/e-		50μV/e-	-	-	65μV/e-	-	-
Frame rate(fps)	90 (HP)	2K-10K (HS)	60	30	15	60	10-60	30
Read noise	407μV	670μV	170μV	-	-	234μV	297μV	320μV
Depth noise	<1.77% @ 0.4~5.4m	<1.82%[3] @ 0.3-1.4m	<0.92% @ 0.4-4m	<0.41% @ 0.6-3.0m	<0.6% @ 1-30m	<0.57% @ 0.4-4.0m	<0.59% @ 0.75-4.0m	<0.5% @ 0.8-4.2m
Chip power	77.8mW	40.1mW[3]	290mW	-	-	160mW	223mW	2.1W
FoM(pJ/pixel)	216	86	9.1	-	-	12.4	131	806
R-FoM(pJ/pixel)	16	18.4	0.91	-	-	1.24	26.8	153.5
Annotation		X8 multi-readout	2x2 binning available	4-pixel interpolation	X3 multi-readout	-	2-pixel interpolation	-

[1]External ADC was used; [2]DC equation for 4-tap SP modulation was used; [3]Measured at 2 Kfps with 33.3ms refresh period;

(BSI) and 3-D stacking technologies. These would improve the layout flexibility and area efficiency to achieve higher fill-factor. Second, IR-responsivity enhancing technologies such as microlens, deep trench isolation (DTI), and pyramid surfaces for diffraction (PSD) structure [28] can be adopted to reduce the exposure time, resulting in a higher frame speed and lower power consumption. Third, to achieve better DC, creating doping gradient on P-epi [24] is beneficial to increasing charge collection efficiency in the vertical direction, resulting in better depth precision. Lastly, the RN can be reduced by implementing a column-ADC circuit and using higher density in-pixel memory with textured deep trench SiN capacitors [29] or high-capacity DRAM capacitors [30], [31]. Moreover, by utilizing high-density capacitor and 3-D stacking techniques, it is possible to implement more in-pixel memory, which can extend the duration of burst imaging even at a higher frame rate. Note that the frame time of HS mode is only limited by MRST and sampling time, which is less than 1 μs in this work.

These advancements in sensor technology can enable higher resolution, lower power consumption, faster frame rate, and better ranging performance for the development of HS and high precision iToF depth imager.

IV. CONCLUSION

An iToF range image sensor with in-pixel analog memory array and sub-frame ToF operation was developed and demonstrated the unprecedented HS range imaging as well as the depth precision enhancement. The fabricated $22.4^H \times 16^V$ μm pitch, $134^H \times 132^V$ 4-tap pixel iToF imager exhibited up to 10 Kfps range imaging with the HS mode and <1.77% depth noise for 0.4–5.4 m range with the HP mode. This development opens a new avenue for HS 3-D imaging applications for machine vision and more.

REFERENCES

[1] Y. Watanabe, "High-speed optical 3D sensing and its applications," *Adv. Opt. Technol.*, vol. 5, nos. 5–6, pp. 367–376, Dec. 2016.

[2] I. Gyongy et al., "High-speed vision with a 3D-stacked SPAD image sensor," *Proc. SPIE*, vol. 11721, Apr. 2021, Art. no. 1172105.
[3] C. S. Bamji et al., "IMpixel 65 nm BSI 320 MHz demodulated TOF image sensor with 3 μm global shutter pixels and analog binning," in *IEEE Int. Solid-State Circuits Conf. (ISSCC) Dig. Tech. Papers*, Feb. 2018, pp. 94–96.
[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 VGA indirect time-of-flight CMOS image sensor with 4-tap 7-μm global-shutter pixel and fixed-pattern phase noise self-compensation," *IEEE J. Solid-State Circuits*, vol. 55, no. 4, pp. 889–897, Apr. 2020.
[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 Circuits*, vol. 55, no. 11, pp. 2849–2865, Nov. 2020.
[7] M. Keel et al., "A 1.2-Mpixel indirect time-of-flight image sensor with 4-Tap 3.5-μm pixels for peak current mitigation and multi-user interference cancellation," *IEEE J. Solid-State Circuits*, vol. 56, no. 11, pp. 3209–3219, Nov. 2021.
[8] J. Kang et al., "An indirect time-of-flight sensor with tetra-pixel architecture calibrating tap mismatch in a single frame," *IEEE Solid-State Circuits Lett.*, vol. 5, pp. 284–287, 2022.
[9] Y. Kato et al., "320 × 240 back-illuminated 10-μM 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.
[10] Y. Kwon et al., "A 2.8 μm 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 *IEDM Tech. Dig.*, Dec. 2020, p. 33.
[11] K. Hatakeyama et al., "A hybrid indirect ToF image sensor for long-range 3D depth measurement under high ambient light conditions," *IEEE J. Solid-State Circuits*, vol. 58, no. 4, pp. 983–992, Apr. 2023.
[12] C. S. Bamji et al., "A 0.13 μm CMOS system-on-chip for a 512 × 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, Jan. 2015.
[13] C. Kuo and R. Kuroda, "A 4-tap CMOS time-of-flight image sensor with in-pixel analog memory array achieving 10 Kfps high-speed range imaging and depth precision enhancement," in *Proc. IEEE Symp. VLSI Technol. Circuits (VLSI Technol. Circuits)*, Jun. 2022, pp. 48–49.
[14] S. Han, T. Takasawa, K. Yasutomi, S. Aoyama, K. Kagawa, and S. Kawahito, "A time-of-flight range image sensor with background canceling lock-in pixels based on lateral electric field charge modulation," *IEEE J. Electron Devices Soc.*, vol. 3, no. 3, pp. 267–275, May 2015.
[15] K. Yamada et al., "A distance measurement method using a time-of-flight CMOS range image sensor with 4-tap output pixels and multiple time-windows," in *Proc. Int. Symp. Electron. Imag. Sci. Technol.*, 2018, pp. 3261–3264.
[16] S. Lee, K. Yasutomi, M. Morita, H. Kawanishi, and S. Kawahito, "A time-of-flight range sensor using four-tap lock-in pixels with high near infrared sensitivity for LiDAR applications," *Sensors*, vol. 20, no. 1, p. 116, Dec. 2019.
[17] T. Sawada, K. Ito, M. Nakayama, and S. Kawahito, "TOF range image sensor using a range-shift technique," in *Proc. IEEE Sensors*, Oct. 2008, pp. 1390–1393.
[18] 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.
[19] M. Suzuki et al., "10 Mfps 960 frames video capturing using a UHS global shutter CMOS image sensor with high density analog memories," *Work*, vol. 10, no. 11, pp. 10–12, 2017.
[20] F. Roy, Y. Cazaux, P. Waltz, P. Malinge, and N. Billon-Pierron, "Low noise global shutter image sensor working in the charge domain," *IEEE Electron Device Lett.*, vol. 40, no. 2, pp. 310–313, Feb. 2019.
[21] K. Miyauchi et al., "Pixel structure with 10 nsec fully charge transfer time for the 20 m frame per second burst CMOS image sensor," *Proc. SPIE*, vol. 9022, Mar. 2014, Art. no. 902203.
[22] K. Kondo et al., "A built-in drift-field PD based 4-tap lock-in pixel for time-of-flight CMOS range image sensors," in *Proc. Extended Abstr. Int. Conf. Solid State Devices Mater.*, Sep. 2018, pp. 9–13.
[23] S. Caranhac and Y. Thenoz, "Driving gate charge coupled device," U.S. Patent 0 649 567, Jul. 18, 2000.
[24] C. Tubert, L. Simony, F. Roy, A. Tournier, L. Pinzelli, and P. Magnan, "High speed dual port pinned-photodiode for time-of-flight imaging," in *Proc. IISW*, 2009, pp. 1–3.
[25] R. H. Nixon et al., "256 × 256 CMOS active pixel sensor camera-on-a-chip," *IEEE J. Solid-State Circuits*, vol. 31, pp. 2046–2050, Dec. 1996.
[26] T. Kasugai et al., "A Time-of-Flight CMOS range image sensor using 4-tap output pixels with lateral-electric-field control," *Electron. Imag.*, vol. 28, no. 12, pp. 1–6, Feb. 2016.
[27] M. Reynolds, J. Dobos, L. Peel, T. Weyrich, and G. J. Brostow, "Capturing time-of-flight data with confidence," in *Proc. CVPR*, 2011, pp. 945–952.
[28] I. Oshiyama et al., "Near-infrared sensitivity enhancement of a back-illuminated complementary metal oxide semiconductor image sensor with a pyramid surface for diffraction structure," in *IEDM Tech. Dig.*, Dec. 2017, p. 16.
[29] K. Saito et al., "High capacitance density highly reliable textured deep trench SiN capacitors toward 3D integration," *Jpn. J. Appl. Phys.*, vol. 60, Mar. 2021, Art. no. SBBC06.
[30] J.-K. Lee et al., "A 2.1e$^-$ temporal noise and -105 dB parasitic light sensitivity backside-illuminated 2.3 μm-pixel voltage-domain global shutter CMOS image sensor using high-capacity DRAM capacitor technology," in *IEEE Int. Solid-State Circuits Conf. (ISSCC) Dig. Tech. Papers*, Feb. 2020, pp. 102–104.
[31] Y. Oh et al., "A 140 dB single-exposure dynamic-range CMOS image sensor with in-pixel DRAM capacitor," in *IEDM Tech. Dig.*, Dec. 2022, p. 37.

Chia-Chi Kuo (Student Member, IEEE) received the B.S. and M.S. degrees in electrical engineering from the National Tsing Hua University, Hsinchu, Taiwan, in 2013 and 2016, respectively. He is currently pursuing the Ph.D. degree with the Graduate School of Engineering, Tohoku University, Sendai, Japan.

From 2016 to 2019, he was with PixArt Imaging Inc., Hsinchu, where he developed analog-mixed-signal circuits for various sensor systems. His current research interests include the high-speed and time-of-flight CMOS image sensors.

Rihito Kuroda (Member, IEEE) received the B.S. degree in electrical engineering and the M.S. and Ph.D. degrees in management science and technology from Tohoku University, Sendai, Japan, in 2005, 2007, and 2010, respectively.

He was a Research Fellow of the Japan Society for the Promotion of Science Research, from 2007 to 2010. Since 2010, he has been with Tohoku University, where he is currently a Professor with the New Industry Creation Hatchery Center, and also with the Graduate School of Engineering. His research interests include advanced process and device technologies and CMOS image sensors and high-precision measurement technologies.

Prof. Kuroda serves as a Committee Member for IEEE IEDM, VLSI Symposium, a Board Member of the International Image Sensor Society, and an Associate Editor for the IEEE TRANSACTIONS ON ELECTRON DEVICES.