Tap mismatch mitigation of 3 µm 2-tap pixels of indirect Time-of-Flight image sensor for high-speed depth mapping

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

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

Abstract This paper presents the development of a VGA-resolution stacked back-illuminated (BI) indirect time-of-flight (iToF) image sensor with 3.0 µm 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°, 180°, 90°, 270°) 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°, 180°, 90°, and 270° for TapA and 180°, 0°, 270°, and 90° 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° against 0° and 270° against 90°) 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 (φ), 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.

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

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

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 µm 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 2×4 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 (Cmod) 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 μm 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.

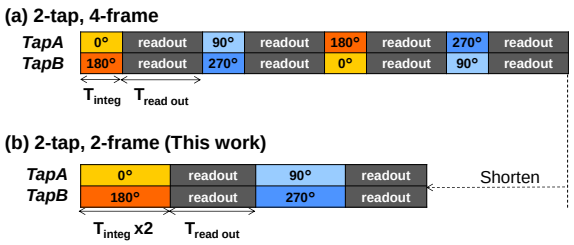

Fig. 1. The conventional 4-frame and proposed 2-frame readout sequences

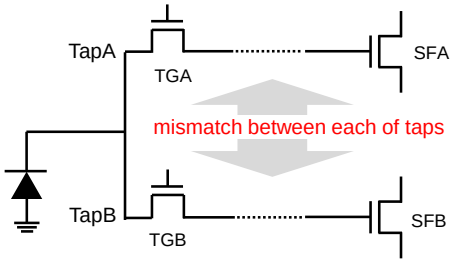

Fig. 2. Tap mismatch of 2-tap iToF pixel

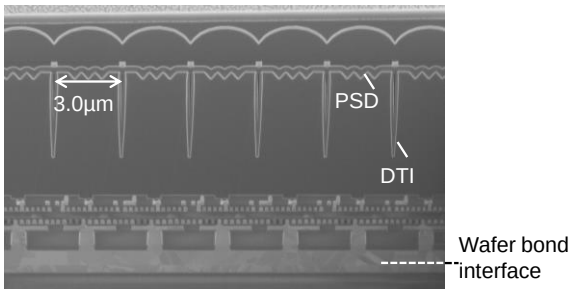

Fig. 3. The Cross Section of our device

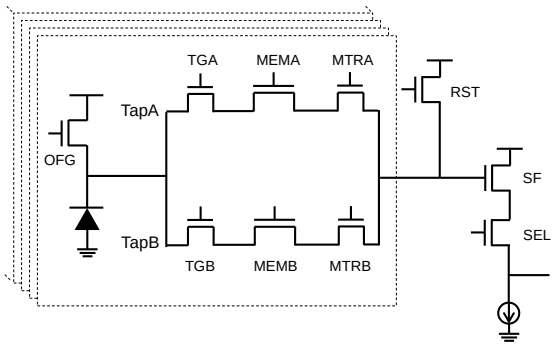

Fig. 4. Pixel architecture

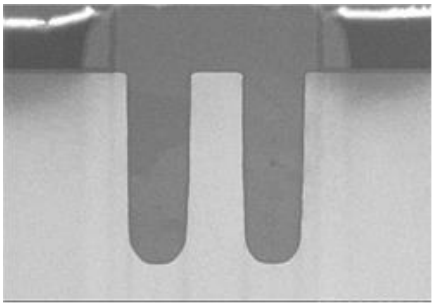

Fig. 5. Cross-section of Dual-VG

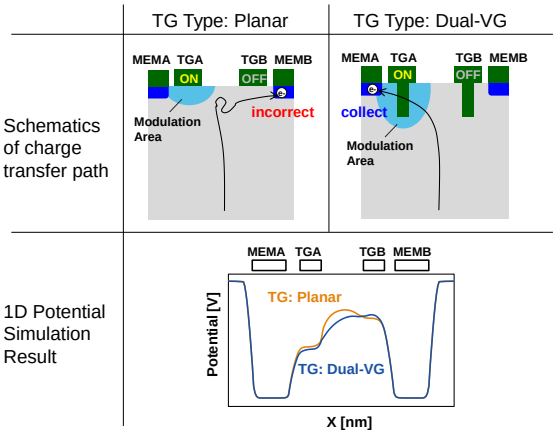

Fig. 6. Comparison between Planar TG and Dual-VG

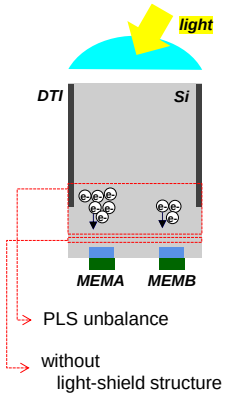

Fig. 7. The source of PLS mismatch

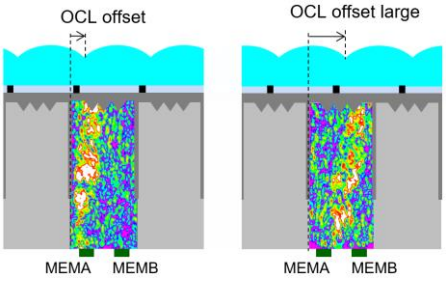

Fig. 8. OCL offset optimization and electric field distribution of 940nm light

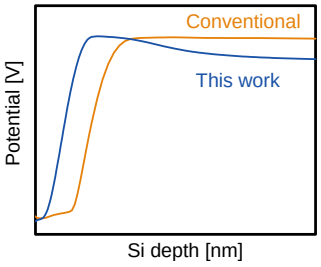

Fig. 9. Simulation results of MEM's vertical potential profile

		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 μm		3.5 μm	4.6 μm	3.5 μm	2.8 μm
	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

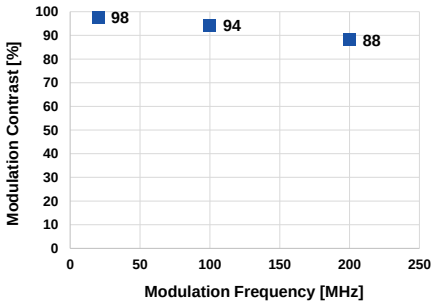

Fig. 10. Modulation Frequency dependency of Modulation Contrast

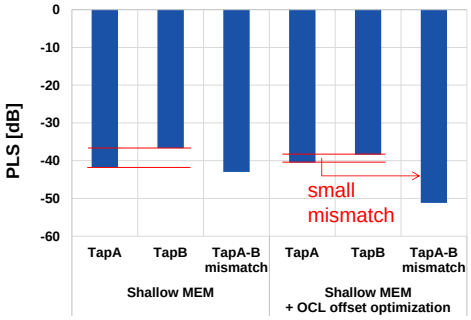

Fig. 11. PLS mismatch

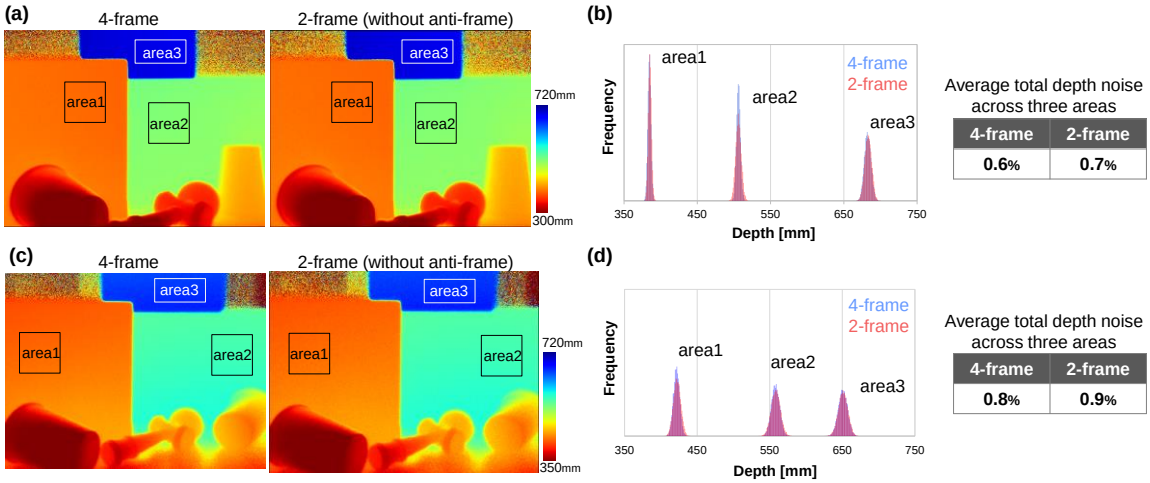

Fig. 12. Single depth images taken at indoor and outdoor (200 MHz frequency, 800 μs total integration time)
(a) depth image at indoor (0.1 klux), (b) histogram of each area at indoor
(c) depth image at outdoor (10 klux), (d) histogram of each area at outdoor

Fig. 13. Group photo taken by the sensor