博 士 学 位 論 文
Doctoral Thesis

論文題目
Thesis Title

A Study on High Precision and High Speed

Time-of-flight CMOS Image Sensor for Range Imaging

(Time-of-flight CMOS イメージセンサによる

高精度・高速距離イメージングに関する研究)

東 北 大 学 大 学 院 工 学 研 究 科

Graduate School of Engineering,
TOHOKU UNIVERSITY

専攻/Department: Management Science and Technology

学籍番号/ ID No: B9TD9801

氏 名 /Ｎａｍｅ: Chia-Chi Kuo

Advising Professor at Tohoku Univ.	Professor Kenji Nakamura
Research Advisor at Tohoku Univ.	
Dissertation Committee Members Name marked with "○" is the Chief Examiner	○ Prof. Kenji Nakamura 1 Prof. Makoto Takahashi 2 Prof. Shuichi Ishida 3 Prof. Tatsuo Yoshinobu 4 Prof. Rihito Kuroda 5 6

A Study on High Precision and High Speed
Time-of-flight CMOS Image Sensor for Range Imaging

ABSTRACT : This dissertation contributes to the field of 3-dimensional (3D) imaging technologies by presenting an in-depth study of a high-precision and high-speed indirect time-of-flight (iToF) CMOS image sensor. The research covers various aspects, including background knowledge, key technologies, system development details, potential applications, and proposed future improvements. By addressing the challenges and limitations of existing iToF imaging systems, this study aims to broaden the potential applications of 3D imagers in the field of machine vision.

In Chapter 1, an overview of 3D imaging technologies and their key requirements is provided. Various approaches, including stereo vision, structured light, and time-of-flight, are discussed. Among these approaches, the iToF ranging methodology is selected for this study due to its potential benefits, such as lower power consumption, good depth precision with a scalable range, small system footprint, and the ability to achieve higher frame rates. The operation principle and important parameters of the iToF ranging system are also described, laying the foundation for the subsequent chapters.

Chapter 2 delves into the key technologies required to realize the targeted iToF imager. These technologies include the adoption of a 4-tap short pulse (SP) modulation, sub-frame ToF operation, high-speed charge modulator, and high-density memory with auto-zeroing operation. The 4-tap iToF scheme is implemented to reduce motion artifacts in scenes with moving objects, while the SP modulation improves depth precision under a high signal-to-noise ratio (SNR) system. The sub-frame ToF operation with a high conversion gain (CG) pixel is proposed to enhance the SNR, with an optimal subframe number of 8 identified for achieving better depth image quality. Charge domain binning is employed to reduce system noise in high-precision (HP) mode, while subframe readout enables high temporal resolution in high-speed (HS) mode. To maintain good demodulation contrast (DC) and increase charge collection efficiency, a relatively large photodiode (PD) with an optimized potential gradient is required. Furthermore, in-pixel memory with high-density Si trench capacitors and auto-zeroing operation enables sub-frame operation with a smaller memory array and higher PD fill-factor (FF).
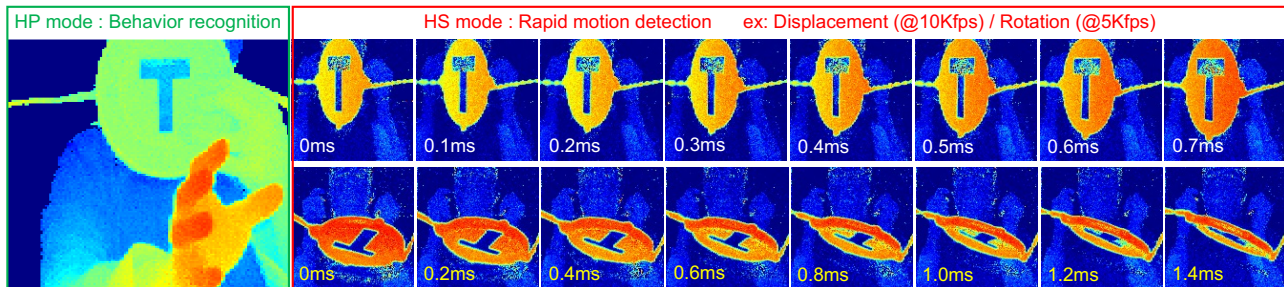
Chapter 3 provides detailed information on the implementation and verification of the prototype iToF sensor with high-precision and high-speed depth imaging capabilities. The proposed iToF imager features a pixel array of 134H×132V pixels, each with a size of 22.4H μm×16V μm. Each pixel is equipped with a 4-tap high-speed charge modulator and a 4×8 memory array. The optimized high-speed charge modulator can collect the photon-generated electrons within 0.8 ns. The sub-frame ToF operation consists of 8 subframes with SP modulation. HP mode imaging adopts 4-tap modulation with memory averaging readout, while HS mode uses pseudo-2-tap operation and obtains 8 frames of burst images by reading out the memory individually. The iToF imager achieves an unprecedented performance in HS mode, achieving up to 10 Kfps range imaging. Additionally, less than 1.77% depth noise and an R-FoM (Range Figure-of-Merit), which determines the best efficient working range, of 16pJ/pixel are achieved at 0.4-5.4 m in HP mode. The HP mode showcases the ability for constructing 3D environments, while the HS mode can be

applied to advanced machine vision tasks such as rapid motion recognition and analysis. The iToF imager shows promise for use in automotive safety systems for behavior monitoring and accident detection, representing a new direction for high-speed 3D imaging applications in machine vision and beyond.

Chapter 4 introduces a novel 2-tap 4-phase (2T-4PH) iToF ranging method using half-pulse (HP) modulation with sub-frame operation. HP modulation is achieved by using a modulated light with half the pulse width and double the amplitude, which allows obtaining different ranging algorithms for HP1 and HP2 with BGLC (Background Light Canceling) ranging results. The operation principle of the ranging method is explained, and theoretical depth noise equations are derived and verified through experiments. Compared to the conventional continuous square-pulse (SP) modulation, the HP modulation exhibited a depth noise reduction of >25% and >29% for the 0.4-3 m range using the HP1 and HP2 methods, respectively, and can be utilized in different scenarios. HP1 is recommended for indoor applications requiring high depth precision and high frame rate, while HP2 has an advantage in lower SNR ranging systems or long-distance ranging under strong ambient light. The combination of the 2T-4PH scheme and sub-frame operation effectively suppresses motion artifacts due to the compact modulation periods, ensuring high-quality depth imaging with moving targets. The proposed technologies can be adopted in 2-tap iToF sensors to enhance their ranging performance without increasing the modulation period or frequency, resulting in higher frame rates and lower power consumption.

In Chapter 5, a BSI 3-D stacked 2-tap iToF pixel was designed, which utilizes all the proposed technologies in this research and enables an improved performance in area efficiency, light sensitivity, depth precision, framerate, and high-speed imaging.

In summary, this study provides a comprehensive investigation into a high-precision and high-speed iToF CMOS image sensor, addressing key challenges and limitations in existing iToF imaging systems. The prototype sensor demonstrates an unprecedented 3D imager that can deliver high-quality depth images while maintaining the ability to capture high-speed images. The developed iToF imager showcases potential applications in various fields, particularly in machine vision and automotive safety systems. Finally, the promising directions for the future improvement are discussed and followed by a practical implementation. This study contributes valuable insights and advancements to the field of high-precision and high-speed 3D imaging technologies, setting the stage for further developments in the future.

Index

Chapter. 1.

Introduction

1.1. Research Background

In recent years, the market of CMOS image sensor (CIS) has expended significantly. For the applications of machine vision, such as object tracking, pattern recognition, and motion detection, image sensors have specific requirements that differ from those used on consumer cameras for human eyes. Machine vision aims to enable automated systems to "see" and process visual information for various industrial and commercial tasks. The key requirements of imager sensors include:

➢ Response linearity: Ensuring consistent and reliable image analysis by directly correlating light intensity with sensor output signal.

➢ Frame rate: Detecting and analyzing rapid scene changes or fast-moving objects for improved system responsiveness.

➢ Spatial resolution: Capturing detailed information for precise measurements, defect identification, and pattern / feature recognition.

These specific features ensure machine vision systems to efficiently analyze visual data and perform various tasks effectively. However, relying solely on a conventional 2-dimensional (2D) imager can result in errors when processing complex environments due to the lack of depth information.

Introduction

To improve the reliability and robustness of machine vision, three-dimensional (3D) imaging provides crucial spatial information that enables machines to better understand and interact with their surroundings. Various methods can be used to achieve 3D imaging, including stereo vision, structured light, and time-of-flight, each with its own benefits and limitations that are suitable for specific applications.

- ➢ Stereo vision, which uses two cameras to reconstruct 3D views by calculating the pixel-wise displacement between images, has been widely adopted in autopilot and autonomous systems due to its system flexibility [1].

- ➢ Structured light works by projecting a known pattern onto an object and recovering its 3D geometry using triangulation algorithms. This technique offers great precision for short-range imaging and has been successfully applied in biometric systems for facial recognition [2].

- ➢ Time-of-flight (ToF) technology has gained significant attention in recent years. It obtains a depth map by measuring the traveling time of light emitted from a sensor system to an object using either a direct (dToF) or indirect (iToF) approach. The dToF method uses a single-photon avalanche diode (SPAD) and has been utilized in light detection and ranging (LiDAR) systems, providing consistent ranging quality over a wide range of distances [3]–[6]. On the other hand, the iToF method shows great potential in applications such as behavior monitoring, gesture recognition, object counting, and augmented/virtual reality (AR/VR), and so on. It offers several advantages, including scalable detection range, low power consumption, cost-effectiveness, and system compactness [7], [8].

1.2. The Key Requirements for 3-D Imaging

1.2.1. Global Shutter

The global shutter function, which controls the exposure timing of all pixel arrays simultaneously, can ensure high-quality images without rolling distortion under a moving scene. For machine vision applications, it is crucial to accurately analyze the state of the captured environment. Furthermore, active illumination is required for both structured light and ToF-based 3D imaging techniques, where good synchronization between the emitter and sensor systems is necessary. Therefore, global shutter is indispensable for the realization of range imagers.

1.2.2. Pixel Count

A higher pixel count provides better image resolution, allowing for more detailed information to be observed. In 3D imaging applications such as medical assistance and biometric security systems, higher image resolution is desired. In recent years, advances in process technology have enabled commercial CIS with pixel pitches of less than 1μm, which can be utilized by stereo vision and structured light approaches.

On the other hand, ToF-based range imagers are also rapidly improved, with dToF using SPAD arrays scaled down to pitches of around 5-10μm [9], [10], and iToF structures achieving even smaller pitches of less than 5μm [11]–[14]. Although increasing the pixel count by shrinking the pixel size can result in better image resolution, it can also reduce photon sensitivity, leading to longer exposure periods due to decreased incident light. Therefore, for the development of high-speed imagers, to achieve shorter frame periods with higher pixel counts, it is more practical to expand the area of the pixel array.

1.2.3. Frame Rate

High frame rate imaging allows a better motion detail owing to higher temporal resolution. The 3-D imaging with a frame rate at standard video rates (30/60 fps) can be adopted to object modeling, gesture recognition and behavior monitoring, whereas a performance level around 1 Kfps is needed for rapid object tracking and dynamic projection mapping usages [15]. 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.

A structured light stereo algorithm as well as a dToF sensor has been reported that achieved such frame rate [15], [16]. However, the GPU acceleration in complex post-processing and high-speed readout requirements not only limit the feasibility toward a higher frame rate and image resolution but also increase the system power consumption and complexity that reduce the portability of the sensor system. In contrast, the iToF based range imagers offer several advantages, such as scalable detection range, low power consumption, cost-effectiveness, and compactness of the system. Nevertheless, no iToF approaches have been reported for high-speed 3D imaging.

1.2.4. Power Consumption

Power consumption is a crucial factor to consider when designing sensor systems, especially for portable or battery-powered applications. The larger the pixel array and higher frame rate, the more power is needed for high-speed readout, which can negatively impact the portability and durability of the system. Additionally, heat dissipation is

another concern, as high temperatures can cause thermal noise, lower linearity, and higher dark current, which all lead to lower image quality. Proper cooling mechanisms are necessary to maintain optimal performance and image quality.

For a 3-D imaging system, stereo vision requires two high-resolution cameras, while structured light algorithms require a continuous projector light source and external computation. Similarly, dToF ranging using SPAD structures requires an additional power system to generate high-level negative voltage for reversed biasing. These additional power demands can become a bottleneck when seeking to achieve higher frame rates, higher pixel counts, and more compact ranging systems.

1.2.5. Detectable Range

The detectable range of a 3-D sensor is determined by its ability to provide reliable and accurate measurements at a given distance. The depth accuracy and precision are two key parameters used to evaluate the ranging performance of a 3-D sensor. Depth accuracy refers to the linearity error in the ranging result compared to the ground truth, while depth precision measures the temporal variation in the depth values over successive frames.

The detectable range of stereo vision is strongly related to the distance between two cameras and the baseline, which can be adjusted for different applications. However, achieving compactness and reducing software complexity remain challenges. Structured light has excellent precision at short range (<1m) [2], but its performance decreases rapidly with increasing distance due to the triangulation algorithm and patterned light projection. The dToF approach is usually chosen for wide range detection applications, providing consistent ranging precision from sub-meter to hundreds of meters. In contrast, for higher precision at short to medium range (<5m), iToF ranging methods have been a popular choice that have been reported to achieve less than 1% depth noise [11].

1.2.6. Outdoor Usage

While applying a 3-D imager to an outdoor environment, changes in the ambient lighting condition might affect the ranging performance. The stereo vision requires enough illuminance from a scene to ensure a good image quality for depth calculation. This suggests that supplementary lighting might be needed at night to maintain the reliability for the applications such as autonomous systems. On the contrary, the other technologies, which use the active illumination, will encounter with performance degradation under a strong ambient light environment. Therefore, considering the spectrum of sunlight, the infrared (IR) light emitter with a wavelength of 850 nm or 940 nm is usually preferred for the silicon-based 3-D sensing system.

Fig. 1-1 summarizes the 3-D imaging technologies.
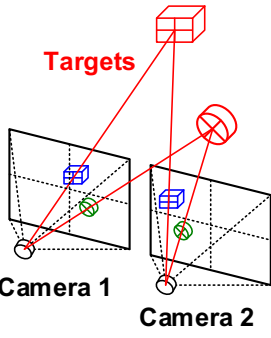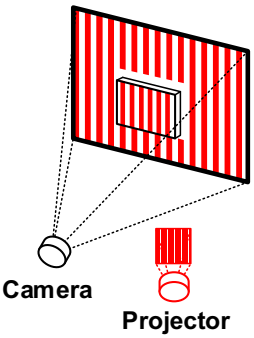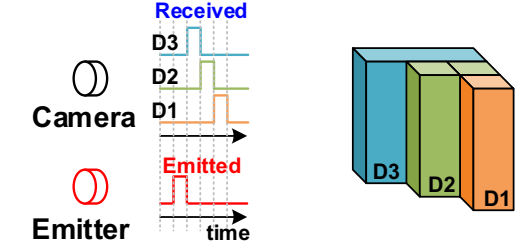
Stereo Vision	Structured Light	Time-of-Flight (ToF)	
		Direct ToF	**Indirect ToF**
		Trigger based	Modulation based
✓ System flexibility ✓ Cost effectiveness ✓ System upgradability △ Software complexity △ System compactness △ Low illuminance scene	✓ Ranging accuracy ✓ System compactness △ Cost effectiveness △ Detection range △ Real-time capacity △ High illuminance scene (Outdoor)	✓ System compactness ✓ Scalable range ✓ Long range detection △ Cost effectiveness △ Image resolution △ Power consumption	✓ System compactness ✓ Power consumption ✓ Cost effectiveness ✓ Scalable range △ High illuminance scene (Outdoor)
Applications			
Autopilot Autonomous system	Biometric Facial recognition	LiDAR system	Behavior monitoring Gesture recognition

Fig. 1-1   3-D imaging technologies

1.2.7. Challenges from 2-D to 3-D Imager

Camera-based 3D imaging technologies, including stereo vision and structured light, to achieve high range resolution and improve the overall performance of these systems, several challenges need to be addressed.

➢ High-resolution cameras: To provide better range resolution, higher resolution cameras are required. However, higher resolution sensors come with larger pixel arrays, which demand higher readout speeds to maintain real-time imaging.

➢ High-speed readout: Increasing the resolution of cameras often leads to larger data volumes. High-speed readout mechanisms and efficient data transfer are essential to handle the increased data flow and maintain high frame rates.

➢ Larger scale imaging system: With the requirement of multi-camera and patterned lighting, the physical size of the imaging system also tends to grow. This may pose challenges in terms of space constraints and portability, especially in applications where compact setups are necessary.

➢ Data processing and computational requirements: Images from high resolution cameras need to be processed and analyzed. Powerful computational resources are necessary to handle the data and extract meaningful information from the images.

Introduction

On the other hand, to achieve high precision in Time-of-flight based 3D imaging, several factors need to be considered:

➢ Sensitivity: The sensor should be designed or optimized to have high sensitivity to the IR wavelength used for the active lighting. This ensures that the sensor can effectively detect the modulated light signals.

➢ Charge Transfer Efficiency: The photodiode's design and material properties should be optimized to maximize the charge transfer efficiency, ensuring that the generated charge is correctly modulated and measured.

➢ Modulation Technique: The method of modulating the light is crucial for accurate ToF measurements. Different techniques, such as square-wave modulation or sinusoidal modulation, can be used to achieve better precision.

➢ Background Noise Reduction: ToF systems are susceptible to background noise, which can affect depth measurements. Advanced noise reduction techniques are employed to minimize the impact of noise on the ranging results.

1.3. Indirect Time-of-Flight (iToF) Technique

Out of the various 3-D imaging technologies discussed earlier, the iToF-based CMOS imagers stand out for their advantages of lower power consumption, good depth precision with scalable range, and small system footprint. Moreover, due to the simplicity of post-processing, these sensors can achieve high frame rates, making them suitable for a wide range of machine vision applications.

1.3.1. Key Parameters

There are several parameters that can influence the performance of an iToF sensor. In this context, the pixel and circuit-dependent coefficients are focused:

➢ **N Taps**

It refers to the number of nodes where the demodulated signals are collected during charge integration in the pixel. Note that this number is not always equal to the number of transfer gates (TG) in some iToF pixel designs which have a charge overflow path.

➢ **Modulation Frequency (f_m)**

With a unit modulation cycle of T_C, the frequency is defined as,

$$f_m = \frac{1}{T_C}$$

Eq. 1-1

Note that for the impulse modulation iToF, the pulse width of modulation light, T_P, is often used for comparison instead of modulation frequency.

➢ **Demodulation Contrast (DC)**

The DC describes the capacity of an iToF pixel to transfer the photon-charges into the designated tap according to a certain modulation frequency. It is defined as,

Introduction

$$DC = \max\left(\frac{2S_n - S_{SUM}}{S_{SUM}}\right), where \ S_{SUM} = \sum_{n=1}^{N} S_n \qquad \text{Eq. 1-2}$$

where N is the number of tap, and the integrated signal in each tap is denoted by S_n. The ideal DC is 100% but will decrease due to the limited charge transfer speed and system bandwidth. In general, smaller pixel size and lower modulation frequency can achieve a higher DC [7], [17].

➢ **Full-Well Capacity (FWC)**

The number of charges that a single tap can store before reaching saturation is known as FWC. Ideally, this capacity should be proportional to the capacitance of the floating diffusion (C_{FD}) and inversely proportional to the conversion gain (CG).

➢ **Dark Random Noise (RN)**

The temporal variation of the consecutive frames under a dark condition that induced by the dark current shot noise and circuit readout noise. It is often expressed in the term of electrons, which represents the total noise power input referred to the floating diffusion (FD). Therefore, it is also inversely proportional to the pixel CG.

➢ **Signal-to-Noise Ratio (SNR)**

SNR is a crucial parameter related to image quality in an iToF system. The signal originates from the reflected modulated light, while the noise consists of read noise (RN) and photon shot noise from the signal and background light (BGL). It can be obtained by,

$$SNR = 20 \ log \frac{N_S}{\sqrt{N_S + N_{BT} + RN^2}} \qquad \text{Eq. 1-3}$$

where N_S and N_{BT} are the number of total integrated electrons in a unit pixel during a single frame from modulated light and BGL, respectively.

1.3.2. Operation Principle

Fig. 1-2 shows the basic idea of an indirect time-of-flight (iToF) ranging operation, where a modulated light is emitted to illuminate a scene. Once the light is reflected by an object, it is received by a photon detector and converted into electrons. Using an iToF pixel, the photon charges are demodulated into individual taps by toggling transfer gates with an opposite phase.

After the modulation, by measuring the relationship between integrated signals in each tap, the time shift ratio (R) and the distance can be obtained by Eq. 1-4, where c is the speed of light and T_P is the pulse width of modulated light.
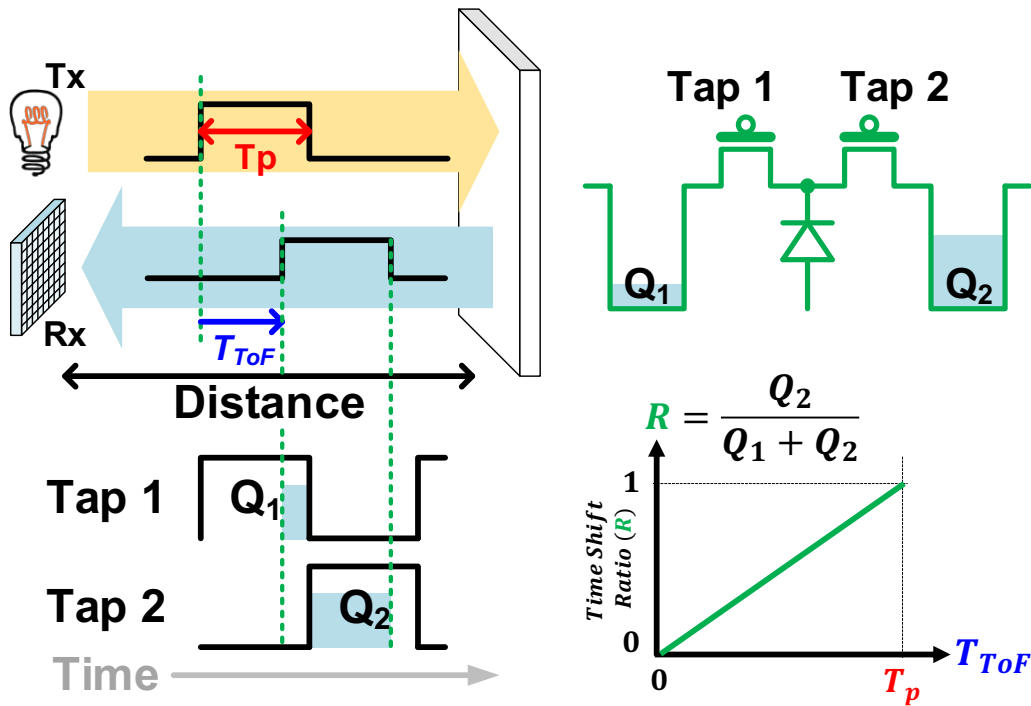

Fig. 1-2 Basic idea of an indirect time-of-flight (iToF) ranging operation.

$$Time\ Shift\ Ratio\ (R) = \frac{Q_2}{Q_1 + Q_2}; \quad Distance\ (d) = \frac{c}{2} \times T_P \times R \qquad \text{Eq. 1-4}$$

In theory, the time shift measurement should be directly proportional to the distance of the object being measured. However, there are various non-ideal factors that can affect the accuracy of the measurement, such as background light, pixel dark current, incomplete charge transfer, limited SNR, clock propagation delay, clock jitter, unstable power supply for the circuit and emitter, reflected stray light, heat buildup during the operation, and even the imperfect shape of the modulated light. As a result, these factors can introduce ranging uncertainty into the iToF sensor system.

1.4. Targets of iToF Sensor Development

To provide a more reliable and robust spatial information with the iToF range imager, lots of research have been conducted in the last decade. The main targets of iToF ranging system development include improving charge transfer efficiency, enhancing depth precision, extending detection range, improving ambient light resistance, and suppressing motion artifacts.

1.4.1. Charge Transfer Efficiency

During the modulation, the light generated electrons are expected to be transferred and collected into the designated tap. However, for the conventional pinned photodiode (PPD) based demodulator, the DC will have a fast degradation as the increase of modulation frequency and pixel pitch due to the weak electric field (E-field) across the photodiode. To improve the ranging performance, the development of a high-speed charge transfer pixel is required.

The types of iToF charge modulator are categorized and described as follows.

Introduction

➢ **Shaped Pinned Photodiode (S-PPD)** [18]

Compares to the conventional square structure PPD, different shapes of PPD with triangular, constant-field and L-shaped have been reported to be useful to increase the DC owing to a stronger lateral E-field. Among these shapes, the L-shaped PPD with a linear doping gradient depicted in **Fig. 1-3**(a), shows a best efficiency that can transfer most of the electrons within 5 ns for a path around 5 μm. In addition, implementing a compact tap structure can make use of the gate-induced fringe field to improve the electron sorting efficiency, achieving a DC of 61% at 100 MHz by TCAD simulation. However, reshaping the PPD scarifies the fill-factor (FF) and results in a lower photon-sensitivity.

➢ **Lateral-Electric-Field Modulator (LEFM)** [19]–[21]

To achieve better control of the E-field, the draining only modulator (DOM) structure with the LEFM has been introduced. As shown in **Fig. 1-3**(b), for a 2-tap LEFM pixel, two sets of gates are placed at both sides of the charge draining path to generate a steeper potential at the depleted region between the two taps. This structure allows effective demodulation of the photon-charge while maintaining an adequate potential barrier for another draining path. The measured DC of ~77% for a 2-tap 16.8 μm pixel was achieved at a gate pulse width of 20 ns. However, the requirement of negative bias and the doubled amount of gates layout may lead to a higher power consumption during demodulation.

Recently, with the shrinking of the iToF pixel, a high lateral E-field can be obtained even without using the DOM structure. A high DC of over 90% was achieved at 100 MHz for a 2-tap 3.5 μm pixel [12].

Introduction

➢ **Current-Assisted Photonic Demodulator (CAPD)** [22], [23]

The CAPD device, as depicted in **Fig. 1-3**(c), creates a strong drift field by implanting the electrode next to each tap instead of using poly-gates to induce the E-field for draining the photon-charges. By applying an alternating voltage biasing, the created current leads the electrons across the diode to the designated taps. This structure can provide high-speed modulation ability even with a larger pixel size. Besides, without the requirement of physical gates, a simpler process and lower cost can be expected. Moreover, without the requirement of physical gates, a simpler process and lower cost can be expected. A DC of around 85% was achieved at 100 MHz for a 2-tap 10 μm pixel. However, the leakage current between the electrodes results in an increase in power consumption while shrinking the pixel size.

➢ **QE modulation detector (QEM)** [24]

The QEM based detector, also known as a photogate modulator, covers the surface of PD region with an extended gate for each tap, as illustrated in **Fig. 1-3** (d). These QEM structures not only enhance the controllability at lateral E-field but also increase the strength of vertical E-field. These QEM structures not only enhance lateral E-field controllability but also increase the strength of the vertical E-field. This indicates that faster charge collection speed can be expected while demodulating the photon-charges, resulting in good DC even with high-frequency operation. It reached up to 78% DC at 320 MHz and 87% at 200 MHz for a 2-tap 3.5 μm pixel [13]. However, the challenge of this structure is the higher parasitic capacitance due to the larger gate size. Therefore, the peak current and propagation delay of the clock driving circuits should be carefully taken care of to prevent inducing fixed pattern noise (FPN) to the depth image [11].
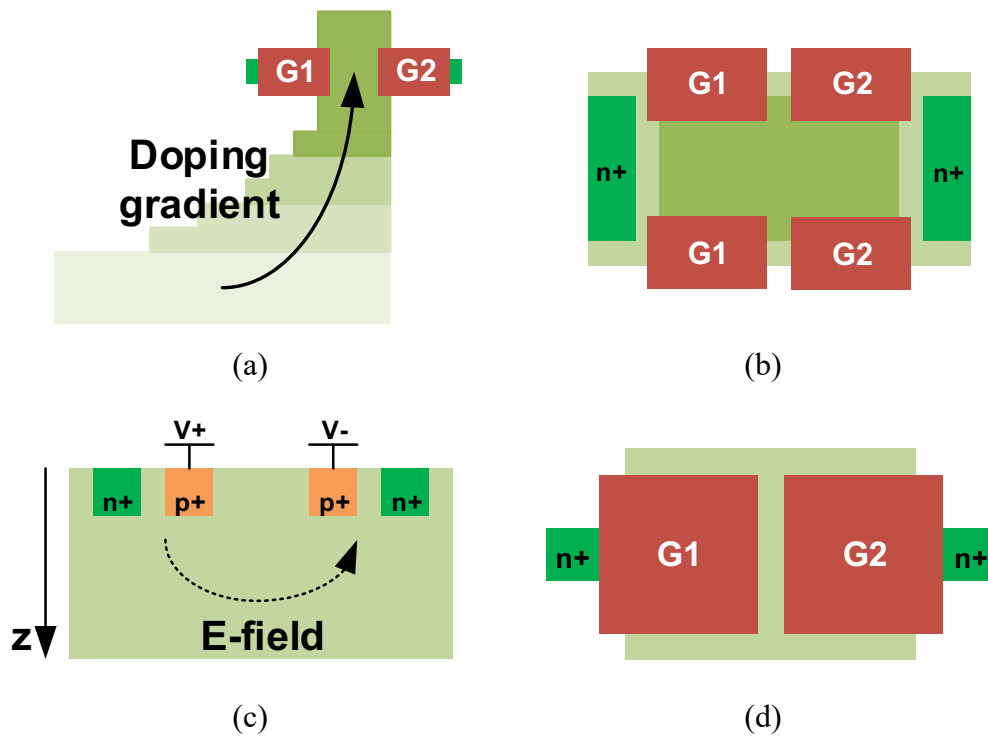
Fig. 1-3 Pixel diagrams of (a) S-PPD (b) LEFM (c) CAPD (d) QEM

1.4.2. Depth Precision Enhancement

As expressed in Eq. 1-5, in an iToF ranging system, the measured depth random noise was revealed to be inversely proportional to the DC, SNR, and modulation frequency [14].

$$Depth\ Noise\ (\sigma_D)\ \propto \frac{1}{DC} \cdot \frac{1}{SNR} \cdot \frac{1}{f_m} \qquad\qquad \text{Eq. 1-5}$$

➢ **Increase the System SNR**

To improve the system SNR of iToF imagers, increasing the FWC and exposure time are common strategies [14], but it comes at the expense of higher input referred noise due to a lower conversion gain (CG) caused by higher FD capacitance. On the other hand, the binning methods can also improve SNR but at a cost of spatial resolution. Methods had been compared among the charge domain, analog domain, and digital domain binning [8]. Among these, the charge domain binning, which integrates electrons from multiple pixels into a shared FD, shows the best SNR improvement efficiency under low photon signal conditions with the same exposure period.

➢ **Increase the Modulation Frequency**

In recent years, there has been a trend towards shrinking pixels to achieve higher image resolution. However, this often leads to lower photon sensitivity and limited FWC, which can result in worse depth noise. To address this, iToF sensors must operate at higher modulation frequencies to recover ranging performance [13]. This presents challenges in both pixel and system implementation, including maintaining DC, mitigating peak current, compensating for clock jitter, ensuring light source robustness, extending detection range, and more.

1.4.3. Detection Range Extension

When adopting an iToF ranging method, it is important to determine the system frame rate, detection range, and cut-off depth performance to meet the requirements of each 3-D imaging application. However, even with a certain emitted light power, which is often limited by eye-safety criteria, the intensity of light reflected by the target will decrease rapidly as a function of distance squared as expressed in Eq. 1-6.

$$p_{target}(d) = \frac{\Gamma \cdot P_{Light}}{\pi(d \cdot tan\,(\theta/2))^2} \qquad \text{Eq. 1-6}$$

where P_{Light} is the power of light source, Γ is the target reflection coefficient, θ is the beam divergence of the emitted light and d is the distance between light source and target. In general, the modulation period will be adjusted to reach the highest possible SNR at the shortest distance based on the FD saturation level. In this case, the lowest depth precision will occur at the maximum detection range due to the degradation of the received signal and SNR.

Although operating with a higher modulation frequency can provide better depth precision, it can also result in a decrease in the maximum unambiguous range (d_{max}) due to the phase wrapping as expressed in Fig. 1-4 and Eq. 1-7,

$$d_{max} = \frac{c}{2 \times f_m} \qquad \text{Eq. 1-7}$$
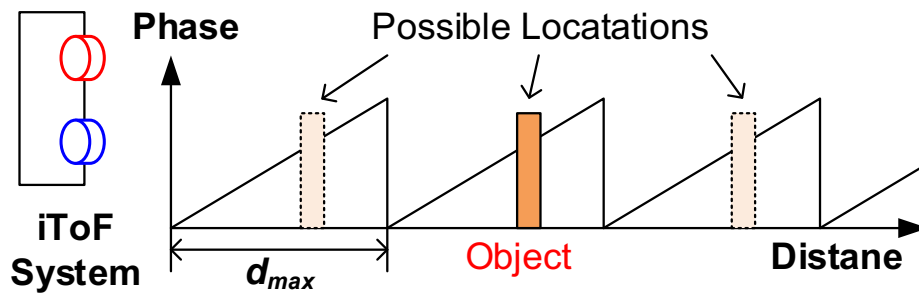

Fig. 1-4 The unambiguous range limitation due to phase wrapping.

The followings describe the concepts to extend the detectable range while able to enhance the depth precision.

➢ **Multi-Frequency Modulation** [25]

As it is already known that higher modulation frequency can reduce ranging uncertainty while lower frequency can allow a wider unambiguous range. Fig. 1-5 shows the combination of these modulations, which can retain depth precision and increase the ambiguous range substantially. The d_{max} is determined by the lowest adopted frequency and the ranging uncertainty is equivalent to the operation with the averaged frequency. This scheme requires a systematic frequency selection to improve the depth precision while ensuring the robustness for phase unwrapping.
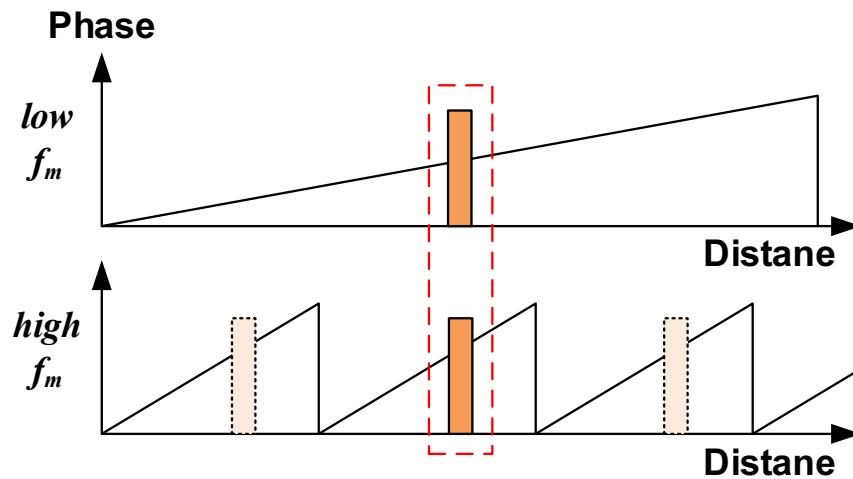

Fig. 1-5 Concept of multi-frequency modulation.

➢ **Range-Shift Operation** [26]

To achieve high range resolution, modulation using a small duty light pulse is useful, but it can result in a short unambiguous range. To address this issue, multiple subframes with range-shifted operation have been reported. As shown in Fig. 1-6, by applying shifted light pulses with a designated offset, each subframe can have an individual scope range. This method offers good flexibility for selecting the

detection range while ensuring depth precision. However, an additional draining gate design on the iToF pixel is required to drain out the uncorrelated light-generated electrons outside the TG demodulation period. Besides, the requirement of more temporal multiplexing can lead to worse motion blur issue when operating with a moving scene.
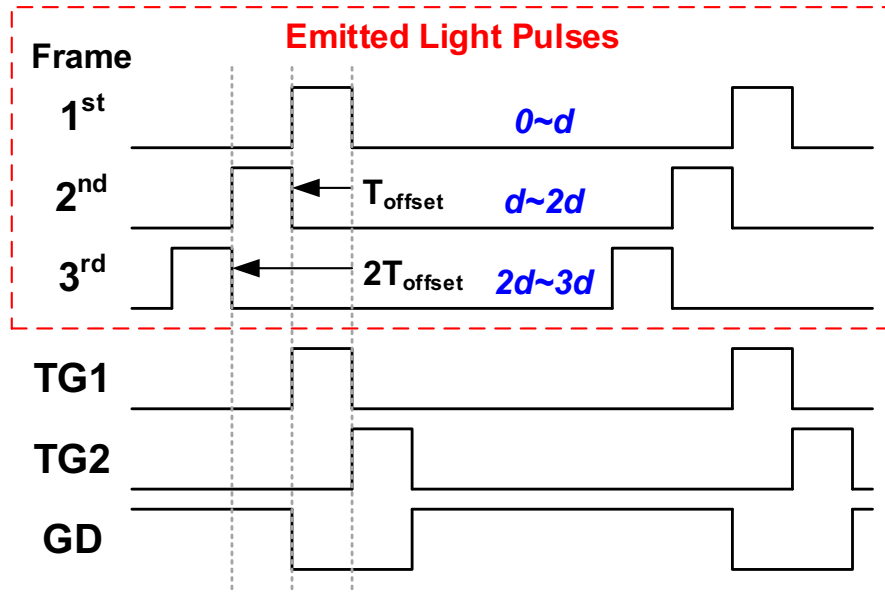

Fig. 1-6 Concept of range-shift operation

1.4.4. Ambient Light Resistance Improvement

To comply with the eye-safety criteria for IR light, the average light intensity during the modulation period is limited. However, the commonly used continuous wave (CW) modulation method experiences a significant degradation in SNR under strong ambient light conditions, resulting in poor depth performance. Although a higher FWC pixel design can enhance the BGL tolerance level, it still involves a trade-off against larger pixel size, higher RN, and longer frame period. Moreover, a BGL cancelling (BGLC) ranging scheme must be adopted to correct the depth error caused by the signal offset generated by BGL. Fig. 1-7 depicts a 2-tap iToF system affected by ambient light.
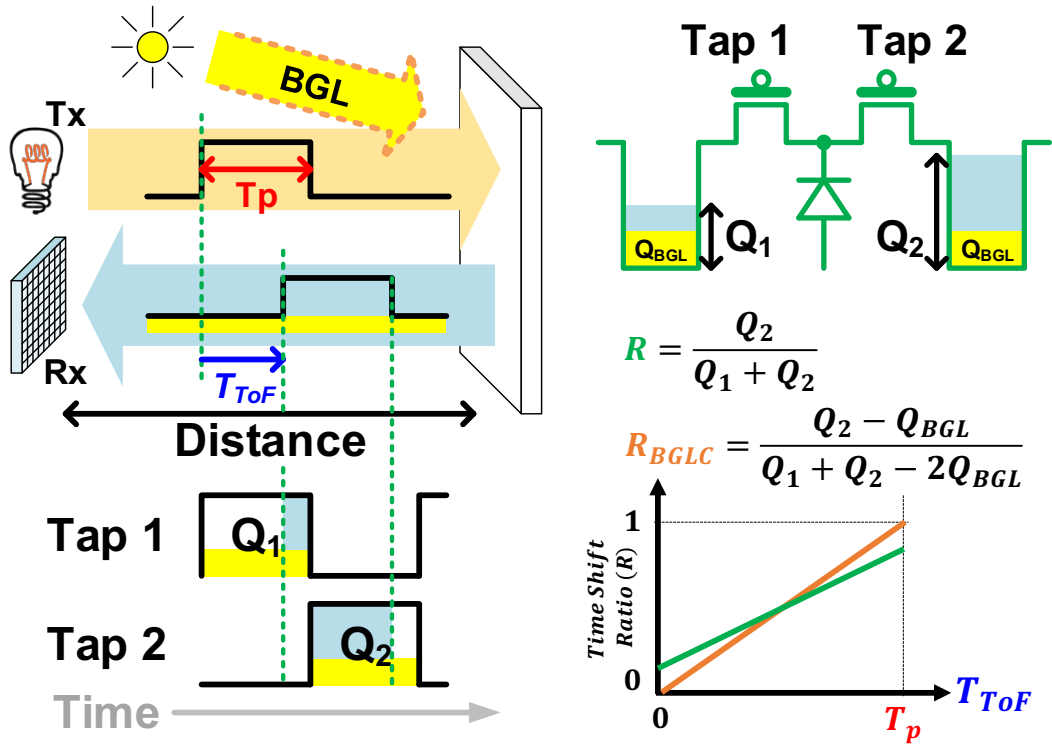
Fig. 1-7 An iToF system affect by ambient light.

$$R = \frac{Q_2}{Q_1 + Q_2}$$

$$R_{BGLC} = \frac{Q_2 - Q_{BGL}}{Q_1 + Q_2 - 2Q_{BGL}}$$

➢ **Short Pulse (SP) Modulation** [21], [27]

Compares to the CW modulation, which modulates the light at 50% duty ratio, the SP modulation uses a smaller duty ratio light pulse but with higher peak power. As shown in Fig. 1-8, the energy-concentrated light pulse can increase the proportion of signal charge to resist the impact from BGL. With the combination of SP modulation and range-shift operation, the outdoor usage ability can be improved. However, the distortion of the light pulse due to the limited bandwidth should be considered.
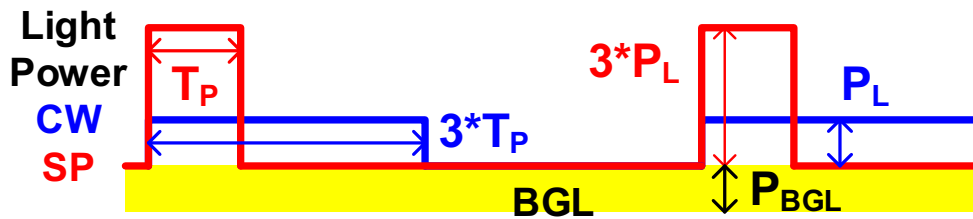

Fig. 1-8 Concept of short pulse (SP) modulation to enhance the BGL resistance.

Introduction

> **On-Chip Background Light Cancelling (BGLC)** [28]–[30]

Due to the requirement of BGLC, the 2-tap iToF needs two modulation periods with a $\pi/2$ phase shift to acquire 4-phase signals, which will be explained in detail in section 4.1. Instead of using the individual signals in Eq. 1-4, the differential charges from both phase are used for the BGLC ranging calculation. In this case, on-chip BGLC scheme can be implemented using either cross-connected capacitors or a charge amplifier with capacitive feedback. Furthermore, the common-mode rejected delta-integration BGLC (Δ-INT BGLC) scheme, as shown in Fig. 1-9, allows the implementation of sub-integration that can accumulate the modulated signal while preventing the FD from saturation to resist stronger BGL in an outdoor environment.
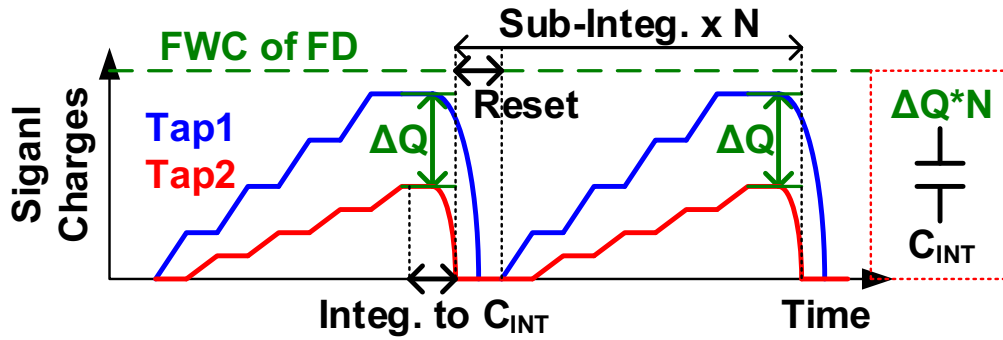
Fig. 1-9 Concept of Δ-INT BGLC scheme

1.4.5. Motion Artifact Suppression

For a conventional 2-tap iToF operation, two consecutive frames are required to recover the depth error induced by ambient light. While capturing moving objects, this temporal multiplexing will lead to an increase of motion blur. Although the high-speed readout circuit can be implemented to reduce the frame-to-frame processing latency, the improvement is still limited especially with a high pixel count imager. To enhance the robustness of 3-D imaging applications, iToF with motion artifact suppression is essential.

➢ Multi-Tap Pixel Structure

The multi-tap architectures implement more than two signal integration nodes for photon-charge storage. An operation example of 3-tap pixel with SP modulation is shown in **Fig. 1-10** (a), where two-tap are used for modulating the signal charge from emitted light, and the other tap samples the offset from BGL [31], [32]. **Fig. 1-10** (b) shows the operation of 4-tap pixel with CW modulation, where the 4-phase signals are sampled within a single modulation cycle [11], [33]. In these cases, the depth can be computed using the signals from a same frame, hence the motion artifact can be significantly suppressed. However, the implementation of multi-tap indicates a lower FF and a larger pixel size. Besides, the offset induces by the tap and readout path mismatch might become a problem for depth non-linearity.

Recently, a pseudo-4-tap (P4T) demodulation method has been reported, where the two adjacent 2-tap pixel from odd-even rows are operated with a $\pi/2$ phase difference [30], [34]. This enables the acquisition of 4-phase signals simultaneously.
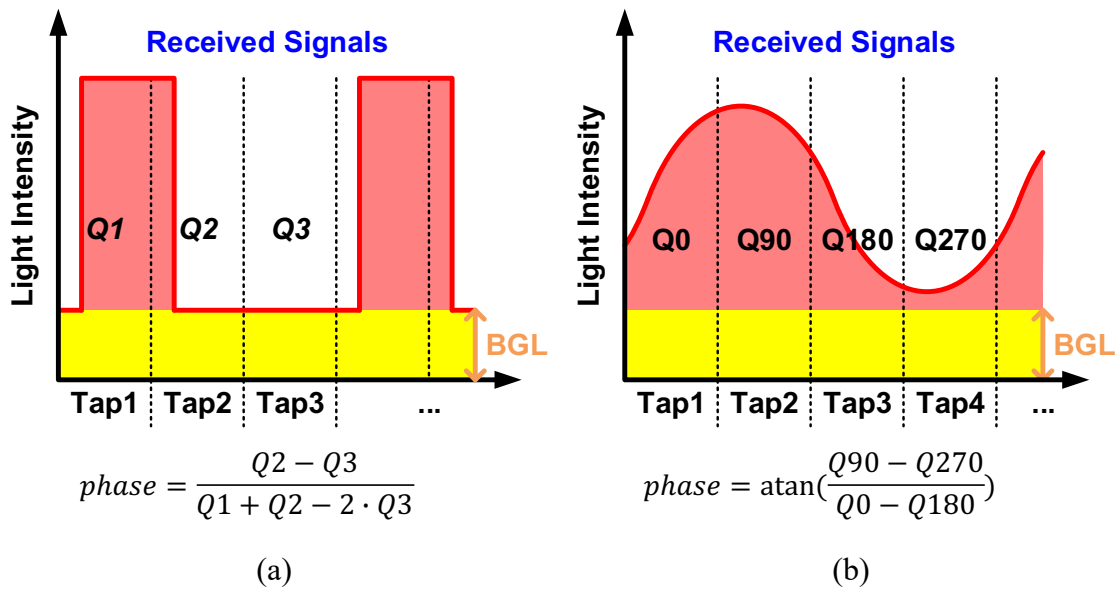

$$phase = \frac{Q2 - Q3}{Q1 + Q2 - 2 \cdot Q3}$$

(a)

$$phase = \operatorname{atan}(\frac{Q90 - Q270}{Q0 - Q180})$$

(b)

Fig. 1-10 Operation concept of (a) 3-tap SP modulation (b) 4-tap CW modulation.

1.5. Purpose of Study

For the development of indirect time-of-flight based CMOS image sensor, research have been reported on the improvement of image resolution, depth precision, detectable range, and outdoor usage ability. These can be helpful for enhancing the robustness and reliability for various existed 3-D imaging applications. However, when it comes to the high-speed 3-D imaging demands, including instant change detection and rapid motion recognition for machine vision purpose, depth images with higher temporal resolution are required. While structured light and dToF approaches have been reported, there are limitations in achieving higher imaging performances and better system portability.

On the other hand, iToF imagers with higher FWC are often preferred to achieve higher system SNR and ambient light tolerance. Additionally, 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. Consequently, current iToF imaging system designs need to compromise between these trade-offs, resulting in a confined field of applications.

In this study, a feasible way to realize an iToF CIS is desired to provide a fine ranging precision while having the capacity to capture high-speed depth images for more potential machine vision applications.

The proposed prototype sensor in this study features a 4-tap high-speed charge modulator, a sub-frame ToF operation utilizing high-density trench capacitors, and an improved ranging method using half-pulse 2-tap 4-phase modulation. The combination of these technologies successfully exhibits the possibility of realizing a high-precision and high-speed 3-D imager. The potential applications and future improvements of this innovation are demonstrated and discussed.

1.6. Contents of Dissertation

Chapter. 1 gave a concise overview of 3-D imaging technologies and their key application requirements. Among these approaches, the indirect time-of-flight (iToF) ranging is chosen for this study due to its comprehensive performance. The operation principle, key parameters, and reported techniques of the iToF sensor were described. The aim of this study is to develop a high-precision and high-speed iToF range image sensor to broaden its potential applications in machine vision field.

Firstly, a prototype high-speed (HS) and high precision (HP) 4-tap iToF range imager is proposed. The concepts of the key technologies, including short pulse modulation, high-speed charge modulator, and sub-frame iToF operation, are explained in Chapter. 2. In Chapter. 3, the implementation details with the design, fabrication, and verification are described, followed by a demonstration and discussion. In addition, a potential application in automotive safety systems is depicted.

Secondly, a practical scheme for future improvement is proposed in Chapter. 4, using a new half-pulse (HP) 2-tap 4-phase ranging method with sub-frame ToF operation. The operation diagram and ranging algorithms are explained, followed by theoretical depth noise calculation. The measured performance is evaluated using the previous prototype iToF sensor, and sample images are demonstrated. Additionally, techniques for further utilization are discussed.

In Chapter. 5, the proposal of next generation high-speed and high precision iToF range imager is described.

Finally, the conclusion of this study is summarized in Chapter. 6.

Chapter. 2.

Key Technologies for High-Precision and High-Speed Range Imaging

2.1. Overview

Fig. 2-1 illustrates the key technologies that will be discussed in detail in this chapter for the realization of a high-precision and high-speed iToF range imager.

They include the utilization of 4-tap short pulse modulation to improve depth noise, the implementation of sub-frame ToF operation to increase system SNR and enable high-speed image acquisition, the integration of a high-speed charge modulator to enhance photon-electron collection efficiency, and the adoption of high-density in-pixel memory with auto-zeroing operation to improve pixel area efficiency.
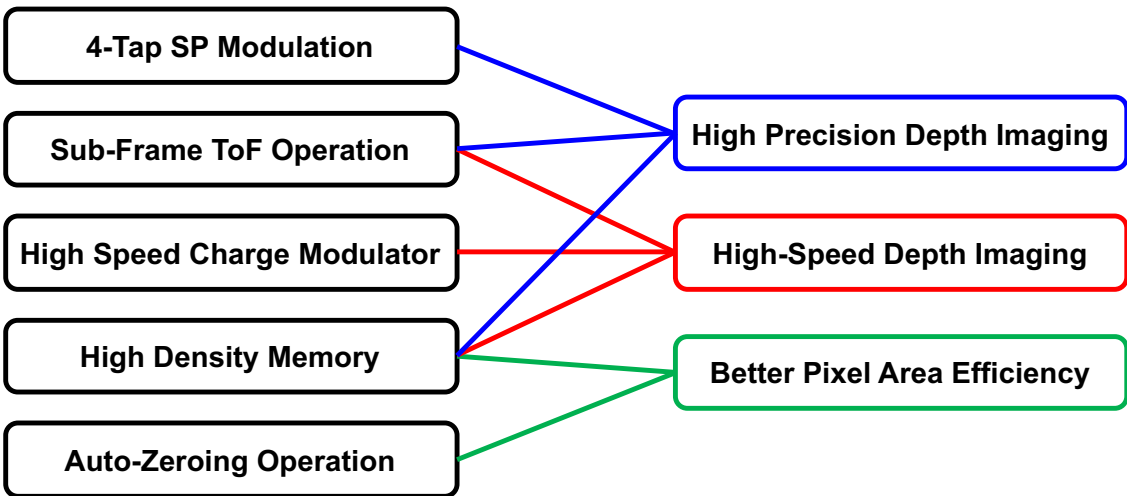

Fig. 2-1 The key technologies in this work.

2.2. 4-Tap Short Pulse (SP) Modulation

While targeting on realizing an iToF imager for high-speed imaging, multi-tap pixel structures are suitable scheme due to the capability of mitigating the motion artifact while capturing a moving scene.

A 4-phase sampling iToF system, utilizing sinusoidal modulated light and a 4-tap charge modulator, has been reported to provide higher depth precision by achieving higher DC with continuous wave (CW) modulation based on the natural sampling theory [33]. In CW modulation, sinusoidal emitted light is preferred owing to its stability and the ability to perform multi-frequency modulation, which extends the unambiguous range. However, ensuring the robustness of phase unwrapping at high modulation frequencies remains a challenge.

On the other hand, short pulse (SP) modulation with 4-tap 1-drain LEFM pixel structures have also been demonstrated for iToF imagers. These structures focus on ambient light resistance and long-range detection. [27]. However, the range-shift operation used in these structures limits the framerate and requires an extra draining gate design in the modulator.
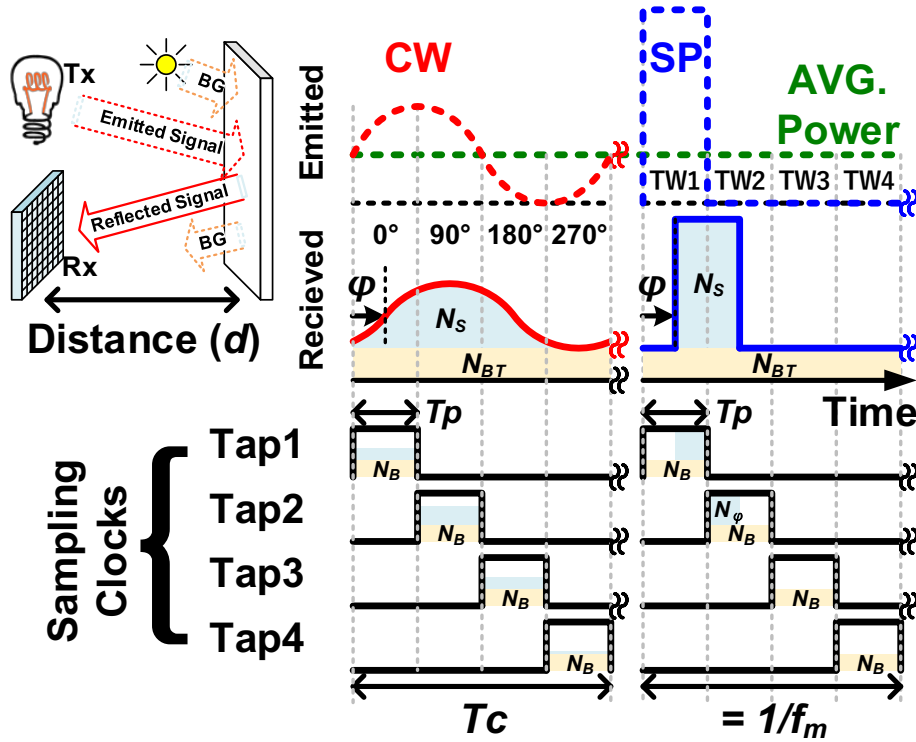
Fig. 2-2 shows the timing diagram of a 4-tap iToF operation using CW and SP modulation, which having an equal averaged light intensity for receiving a same amount of photon electrons from emitted light. The modulation cycle (T_C) is equally divided by 4 sampling clocks with a pulse width of T_P. The reflected signals from the modulation and ambient light are both converted into electrons and demodulated into floating diffusions by Tap1, Tap2, Tap3, and Tap4, and denoted by Q_1, Q_2, Q_3, and Q_4, respectively. The distance d_{CW} and d_{SP} with background light cancelling (BGLC) can be obtained by Eq. 2-1 and Eq. 2-2, assuming the ideal sinusoidal and square pulse modulated lights are given,

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

$$d_{SP} = \frac{c}{2} \cdot \frac{1}{4f_m} \cdot \frac{Q_2 - Q_4}{Q_1 + Q_2 - 2Q_4} \qquad \text{Eq. 2-2}$$
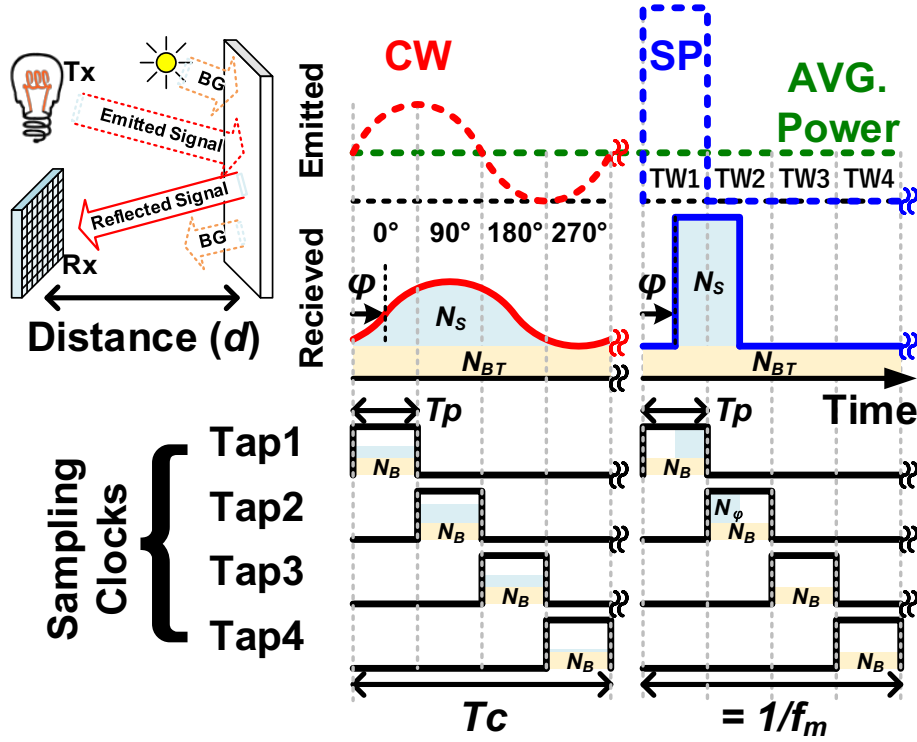
Fig. 2-2 Timing diagram of 4-tap iToF operation

The error propagation method can be adopted for an iToF system to evaluate the ranging result effected by variable's uncertainty [7],

$$\sigma_d = \sqrt{\sum_{n=1}^{N} \left(\frac{\partial d}{\partial Q_n}\right)^2 \cdot \sigma Q_n^{\;2}} \qquad \text{Eq. 2-3}$$

where N is the number of tap and $\sigma Q_n^{\;2}$ is the total noise power including shot noise and readout noise in a tap. Thus, the theoretical depth noise, σ_{d_CW} and σ_{d_SP}, can be derived as follows [21], [33],

$$\sigma_{d\_CW} = \frac{c}{4\pi f_m} \cdot \frac{\sqrt{2} \cdot \sqrt{N_S + N_{BT} + RN^2}}{DC \cdot N_S} \qquad \text{Eq. 2-4}$$

$$\sigma_{d\_SP} = \frac{c}{8 f_m} \cdot \frac{\sqrt{N_\varphi(1 - R_S) + 2(N_B + RN^2)(1 - 3R_S + 3R_S^2)}}{DC \cdot N_S} \qquad \text{Eq. 2-5}$$

where DC is the demodulation contrast and $R_S = N_\varphi/N_S$. The number of total electrons collected in a unit pixel, which was generated by modulated light, is denoted by N_S. The number of electrons generated by BGL in a unit tap is denoted by N_B, and the total electrons is denoted by N_{BT}. The FD input referred readout noise, which came from the pixel transistors and readout circuits, is denoted by RN. With SP modulation, the number of signal electrons in Tap2 is proportional to the phase shift (φ) and is denoted by N_φ.

To compare the performance of CW and SP modulation, different RN levels were applied. The theoretical curves and system SNR are shown in Fig. 2-3. Using a modulation frequency (f_m) of 25 MHz, the unambiguous range spans from 0.4-6m with a pulse duration (T_p) of 10ns and cycle time (T_c) of 40ns. Here, N_S of 20K at the 0.4m without BGL ($N_{BT} = 0$) was assumed. The DC of 90% for both modulation types are used. The SNR is calculated by Eq. 2-6.

$$SNR = 20\log\frac{N_S}{\sqrt{N_S + N_{BT} + RN^2}} \qquad \text{Eq. 2-6}$$
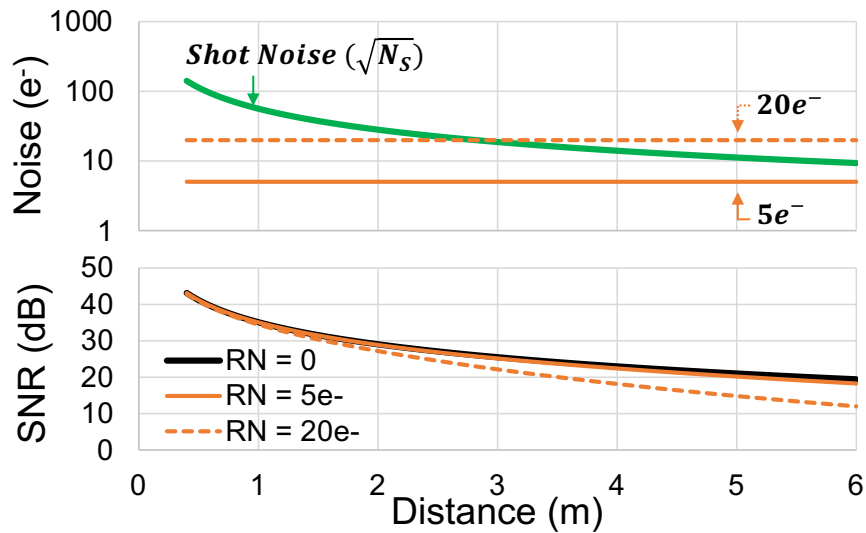
(a)

(b)

**Fig. 2-3 (a) System noise and SNR (b) Theoretical depth noise
over distance with different RN levels.**

As indicated in the simulation result, if the system SNR is predominantly affected by the shot noise of the modulated light, SP modulation can lead to superior depth noise. However, in cases of higher RN, the SNR deteriorates more quickly, leading to a reduced advantage with SP modulation.
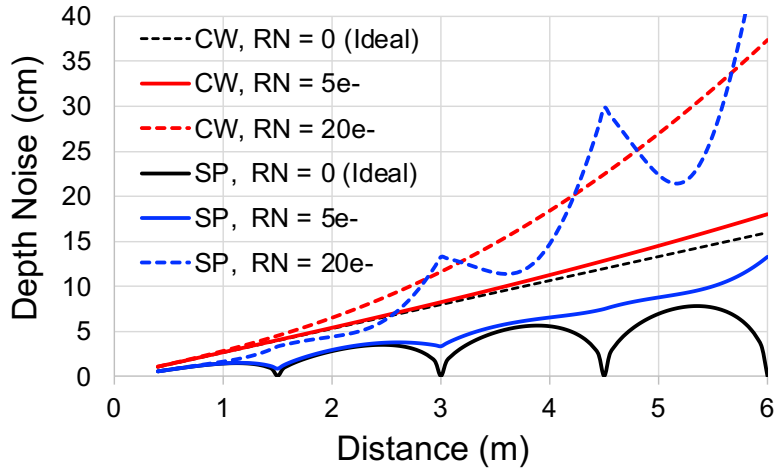
Therefore, to achieve lower depth noise and higher frame rates, the strategy is to develop a 4-tap iToF image sensor using SP modulation with enhanced SNR. In this work, a single frequency SP iToF system without drain gate and range-shift technique is employed to minimize the image processing effort and maximize the frame speed.

In Fig. 2-3, the wavy depth noise curve is shown using the SP modulation. This result occurs because the phase (φ) is repeated within the range from 0 to 1 for the four modulation time windows (TW1~4). Therefore, the signal electron in the Tap (N_φ), which is proportional to φ, will also repeat from 0 to N_S, leading to the wavy noise curve.

2.3. Sub-Frame ToF Operation

In an image sensor system, increasing the full well capacity (FWC) and exposure time are common strategies to achieve higher SNR. However, a higher floating diffusion capacitance leads to a lower conversion gain (CG), which in turn results in higher input referred noise. On the other hand, Pixel binning techniques, including charge domain, analog domain, and digital domain binning, can reduce the readout noise but have the disadvantage of lower spatial resolution [8]. Furthermore, these approaches require longer frame-to-frame latency, increasing the pixel counts due to the array readout time, making them unsuitable for improving higher framerates. Therefore, alternative methods must be explored to achieve both high SNR and high temporal resolution.

In this study, a sub-frame ToF operation is proposed to achieve sufficient SNR while maintaining high temporal resolution imaging. This approach utilizes the memory array by dividing the full modulation period from a conventional ToF operation into subframes, where signals are sampled into separate memory cells after the modulation is completed, as illustrated in Fig. 2-4.

Note that the sub-frame ToF operation can be implemented using two different memory array approaches, depending on the specific requirements. In-pixel memory arrays offer advantages such as shorter sampling periods and lower power consumption, while column-wise memory arrays may be preferred for the demand of pixel shrinkage.
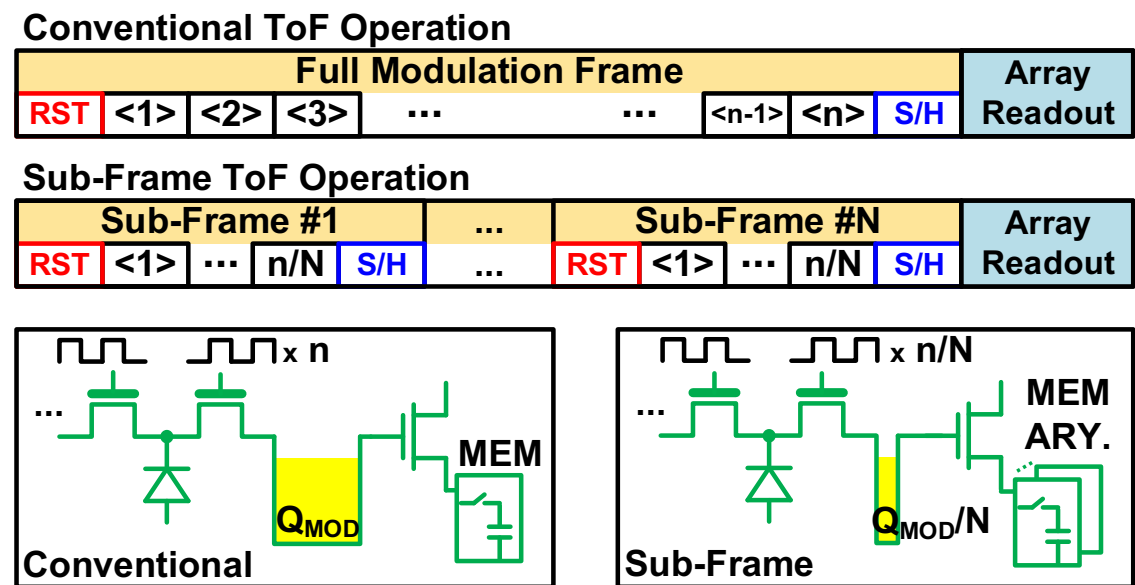
Fig. 2-4 Operation principle of the sub-frame ToF operation.

Through the sub-frame ToF operation, it is possible to perform two types of imaging modes by applying different readout timings, as shown in Fig. 2-5.
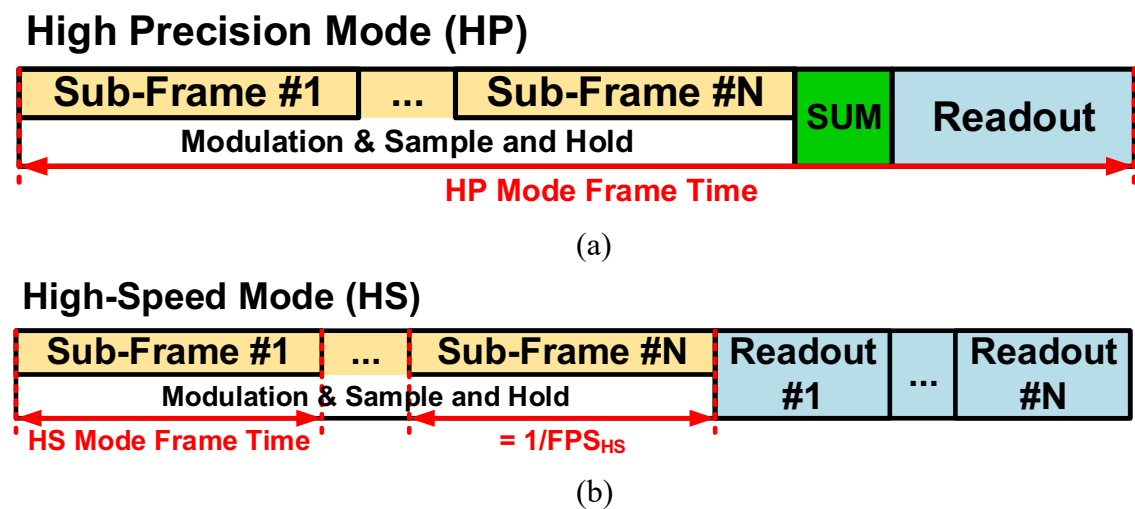

Fig. 2-5 (a) High-precision mode and (b) High-speed mode readout.

2.3.1. High Precision Mode

In HP mode, as shown in Fig. 2-5 (a), the signals sampled from the modulated subframes are combined using charge-domain binning, which is performed by mixing the

signal charges in the memories. This approach helps in reducing the noise from the pixel operation, including shot noise, flicker/thermal noise of pixel transistors, and kTC noise of the in-pixel memory, due to the averaging effect. The theoretical SNR is expressed by Eq. 2-7,

$$SNR_{HP} = 20\log\frac{N_S}{\sqrt{(N_S + N_{PIX}{}^2 + N_{MEM}{}^2)/N_{SubF} + N_{CKTs}{}^2}}$$

Eq. 2-7

$$\left(N_{PIX} = \frac{\sigma_{PIX}}{CG}; N_{MEM} = \frac{\sigma_{MEM}}{CG}; N_{CKTs} = \frac{\sigma_{CKTs}}{CG}\right)$$

where N_{SubF} is the number of subframes, CG is the conversion gain. N_S is the received signal electrons and N_{PIX}, N_{MEM} and N_{CKTs} represent the number of noise electrons from pixel, in-pixel memory, and readout circuits, respectively, which can be calculated using the FD referred noise voltage, σ_{PIX}, σ_{MEM} and σ_{CKTs} divided by CG.
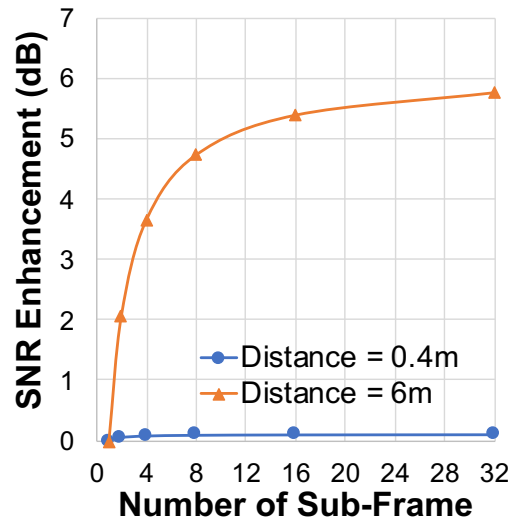
The calculated SNR characteristics using different number of subframes are shown in Fig. 2-6 (a). Firstly, a reference is established using the conventional operation with a CG of $10\,\mu V/e^-$ for comparison with the sub-frame operation. Secondly, a constant CG/N_{SubF} $(\mu V/e^-)$ of 10 and $N_S \cdot N_{SubF}$ (e^-) of 100,000 are used to have an equal amount of signal electrons under a certain exposure condition. In this simulation, the in-pixel memory capacitance of 30 fF is used, and σ_{PIX} of 300μV, σ_{MEM} of 370 μV and σ_{CKTs} of 200 μV are assumed, which are reasonable values for state-of-art image sensors.

Fig. 2-6 (b) illustrates the improvement in SNR achieved by HP mode. For shorter distances, only a slight improvement is observed due to the dominance of shot noise. As the distance increases, a better SNR can be obtained by using a higher number of subframes with a higher CG, leading to a reduction in equivalent noise electrons.

In this study, the optimal number of subframes was found to be 8, taking into account the trade-off between SNR enhancement and hardware complexity.
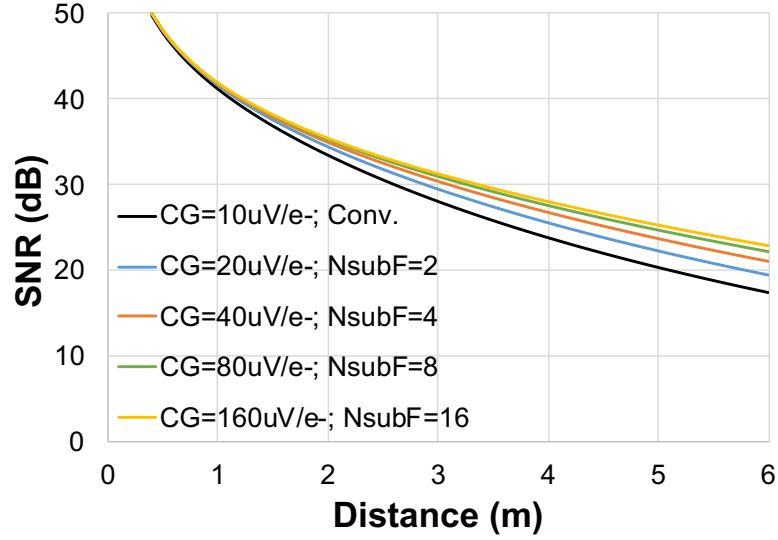
(a)

(b)

Fig. 2-6 Calculated results with different numbers of subframes. (a) SNR characteristic over distance. (b) SNR enhancement at short/long distance.

2.3.2. High Speed Mode

As shown in Fig. 2-5 (b), in HS mode, the signals are read out individually from each memory cell without charge averaging. This allows the acquirement of high-speed

burst imaging with a high temporal resolution. The frame rate in burst mode is determined by the subframe period, which is the sum of the reset, modulation, and sampling periods, as shown in Fig. 2-4. Therefore, to achieve higher frame rates, it is important to increase the photon sensitivity and reduce the modulation counts. The SNR is expressed by Eq. 2-8,

$$SNR_{HS} = 20 \log \frac{N_S}{\sqrt{N_S + N_{PIX}^2 + N_{MEM}^2 + N_{CKTs}^2}} \qquad \text{Eq. 2-8}$$

On the other hand, the available record length is determined by the number of subframes. Thus, the duration of the record time can be expressed by Eq. 2-9,

$$Record\ Duration\ = \frac{N_{SubF}}{framerate} \qquad \text{Eq. 2-9}$$

For the HS mode, increasing the number of subframes allows for a longer recording duration. However, a larger in-pixel memory array will limit the ability of pixel shrinkage for higher image resolution. Therefore, the selection of the subframe number and the desired framerate should be based on the specific application requirements.

Note that the burst images obtained by HS mode can be readout and refresh with a period of $T_{refresh}$, which will be used for comparison in the measurement results. In general, the $T_{refresh}$ is proportional to the pixel count and sub-frame number.

2.4. High-Speed Charge Modulator

Due to the limitation of eye-safety criteria, instead of increasing the optical power, enhancing the photon sensitivity becomes the key to reduce the sub-frame time. Therefore, large photodiode (PD) size with lateral electric field (E-field) optimization [35] allows a better charge collection efficiency while ensuring a good DC.

The followings describe the concepts for optimizing the charge modulator for high-speed iToF imaging:

➢ Increasing the size of PD to increase the number of generated photon charges.

➢ Enlarging the n-layer region to enhance the charge collection efficiency.

➢ Creating a dopant concentration ladder to generate an electric field toward the center of modulator [36]

➢ Implementing a compact 4-tap structure to improve the electron sorting efficiency by utilizing fringe electric field.

➢ Generating a linear potential gradient along the charge transfer path to ensure efficient driving of generated electrons regardless of their initial position.

The implementation and simulation details are to be shown in Section. 3.1.2.

2.5. High Density Memory with Auto-Zeroing Operation

To reduce the sampling period and minimize the exposure deadtime between subframes, an implementation of in-pixel memory array is used in this work. Therefore, a high-density memory is desired to enhance the area efficiency for higher fill-factor (FF).

A high-density Si trench capacitor device has been reported that achieves 25 fF/μm² with low leakage current for signal sample/hold (S/H) usages [37]. However, to enable noise reduction through correlated double sampling (CDS), twice the number of memory cells is required. To address this issue, the auto-zeroing operation is adopted [38].

(1) Floating diffusion (FD), coupling capacitor (CAZ), and memory (CMEM) are reset.

(2) The pixel reset is turned off while the memory reset remains connected.

(3) The memory reset is turned off to initiate the voltage coupling, allowing the integrated signal at FD to be sampled.

Through this process, the generated thermal noise and voltage offset from FD reset can be eliminated. Compares to the conventional CDS operation, the required memory cells are reduced by half, resulting in better pixel area efficiency.

(1)
$$V_{FD} = V_P$$
$$V_{MEM} = V_M$$

(2)
$$V'_{FD} = V_P + V_{OFF} + \tilde{v}_{FD\_kTC}$$
$$V'_{MEM} = V_M$$

(3)
$$V''_{FD} = V'_{FD} + V_{INT}$$
$$V''_{MEM} = V_M + \tilde{v}_{MEM\_kTC} + r \cdot V_{INT}$$
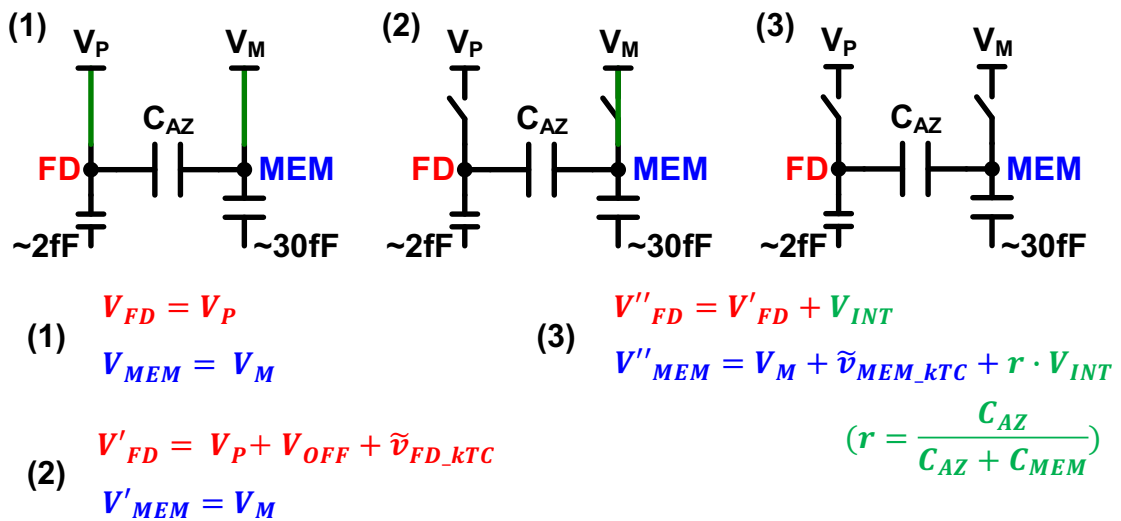$$\left(r = \frac{C_{AZ}}{C_{AZ} + C_{MEM}} \right)$$

Fig. 2-7 The operation principle of auto-zeroing sampling.

2.6. Summary of Chapter. 2

In this chapter, the key technologies for realizing a high-precision and high-speed iToF range imager were introduced.

When capturing moving objects, the 4-tap iToF schemes are useful in mitigating motion artifacts. Meanwhile, using short pulse (SP) modulation with enhanced SNR can lead to better depth noise performance. However, the conventional high FWC strategy suffers from higher FD input referred noise due to lower conversion gain (CG). Besides, the framerate is limited due to the requirement of long exposure time.

Therefore, sub-frame ToF operation with high CG was introduced. In high-precision mode (HP), the system SNR is improved by performing the charge domain binning, with an optimal subframe number of 8 based on SNR enhancement and hardware complexity. On the other hand, in high-speed mode (HS), the high temporal resolution images are obtained by readout the subframes individually.

To reduce the subframe period for high-speed imaging, using a modulator with a larger photodiode (PD) region can increase photon sensitivity. In this case, optimizing the potential gradient to improve electron transfer efficiency is important to maintain good demodulation contrast (DC).

Finally, for implementing sub-frame operation, in-pixel memory is adopted to reduce the sampling period and minimize exposure dead time between subframes. However, the conventional CDS operation requires more memories that limits the pixel fill-factor (FF). Therefore, the combination of high-density capacitor and auto-zeroing operation achieves circuit noise reduction while improving the pixel area efficiency.

Chapter. 3.

Implementation and Verification of the Prototype 4-Tap iToF Range Imager

3.1. Development

This section describes the development of a prototype 4-tap iToF sensor based on the technologies introduced in Chapter. 2.

3.1.1. Chip Architecture

The functional block diagram of the proposed iToF imager is illustrated in Fig. 3-1. The pixel array is composed of $134^H \times 132^V$ pixels arranged in a back-to-back group to enhance the memory array area efficiency, with a pixel size of 22.4^H μm$\times 16^V$ μm [39].

The global drivers are used to generate the control signals for pixel and memory during global shutter operation. The clock tree circuits are utilized to synchronize the trigger of modulation pulse generators, which are implemented using shift registers, across the pixel array. To ensure accurate demodulation under high-frequency modulation, two sets of drivers from both the top and bottom sides of the pixel array are used to drive the demodulation clocks (TGs), minimizing pulse distortion. Additionally, LED trigger and delay control circuits are employed to fine-tune the modulation light pulse edge.

During the rolling readout, where the control clocks are generated by vertical shift register (VSR), the signals in the pixel memory arrays (MEM) are sampled by the column

sample/hold (CSH) with a column pitch of 5.6 µm. Subsequently, the analog output

signals, which are selected by the horizontal shift register (HSR), from the output buffer
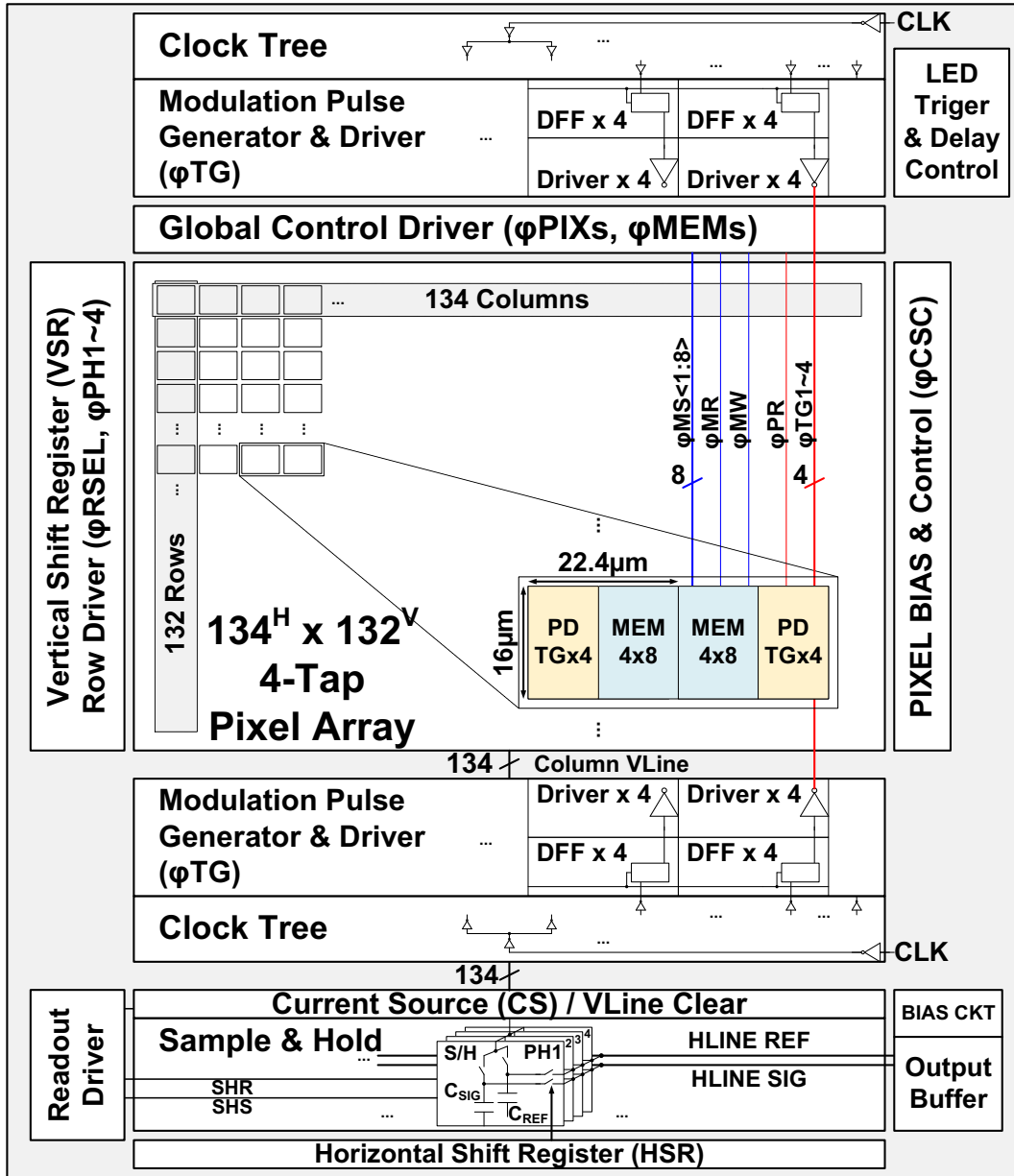
(OB) are quantized by an off-chip 14-bit ADC.

Fig. 3-1 Functional block diagram of the prototype iToF sensor

3.1.2. Pixel Structure

In this prototype sensor, a 4-tap iToF pixel with an in-pixel memory array was developed. The demodulator part includes a high-speed charge collection photodiode (PD), demodulation gates (TG1-TG4), four sets of buried channel source follower (PSF) and current source (PCS) with a cascode switch (CSC). The memory part consists of four auto-zeroing capacitors (C_{AZ}), 4×8 1-T 1-C analog memory array, and control devices (MW, MRST, MPHs), which share a voltage buffer (MSF, RSEL) for column readout. The analog memory capacitor (C_{MEM}) and C_{AZ} are constructed using high-density Si trench capacitors, which provide a capacitance of 30 fF and 170 fF, respectively. The circuit and layout diagrams of the pixel can be seen in Fig. 3-2 (a) and (b).

Fig. 3-2 (c) shows the cross-sectional diagram of the PD. In this work, three levels of n-layer (N1-N3) with different concentration and energy are used to generate the potential ladder to collect the photon-electrons to the center of modulator. Additionally, a spiky triangular-shaped layout was utilized to create a linear potential gradient along the charge transfer path [40], [41]. This ensures that the generated electrons can be driven efficiently regardless of their initial position.
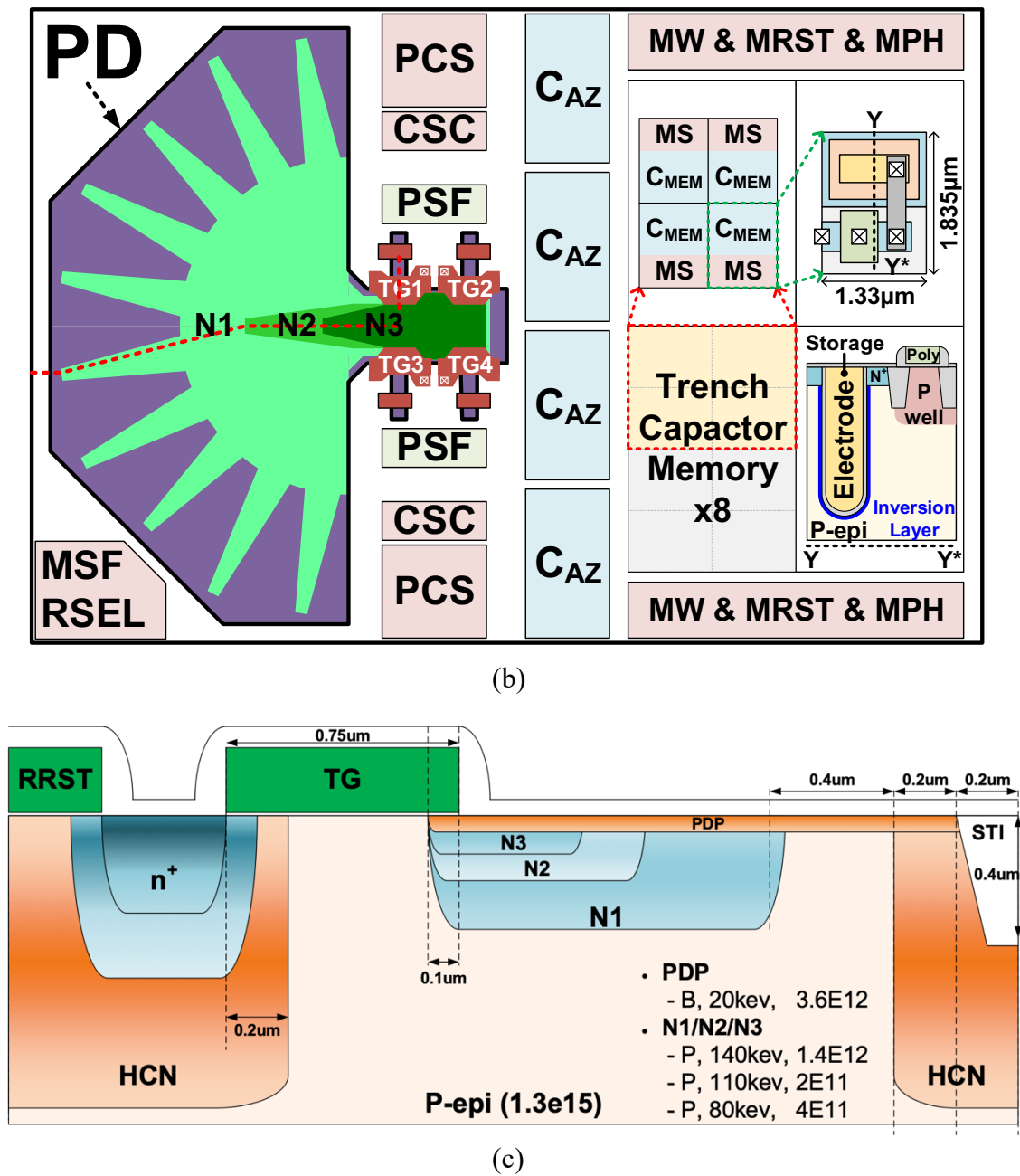

(a)

(b)

(c)
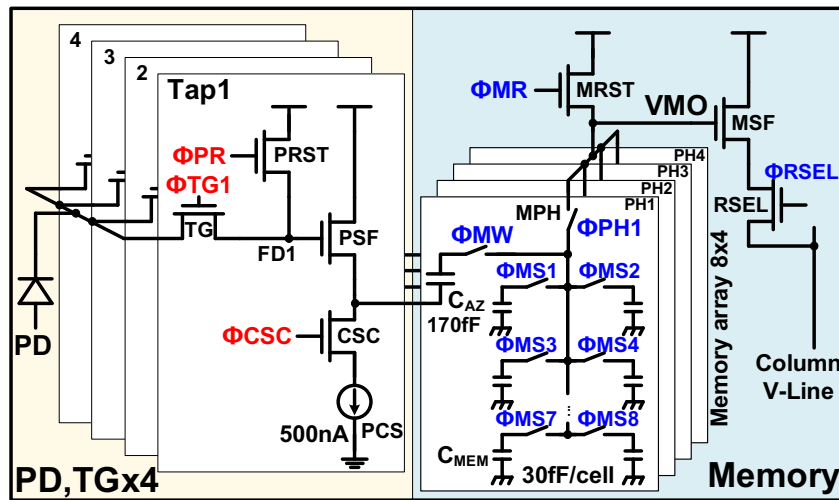
Fig. 3-2 Developed 4-tap iToF pixel with in-pixel memory array

(a) Circuit diagram (b) Layout diagram (c) Cross sectional diagram

The potential diagrams in Fig. 3-3 (a) and (b) show the simulation results of the electron transportation with the demodulation of TG1 and TG2, respectively. In these simulations, the electrons were initially placed at a depth of 1 μm from the far end of the PD. The results demonstrate the successful collection of electrons to FD1 and FD2.

Fig. 3-4 (a) shows that the proposed high-speed charge modulator can collect charges from the photodiode with an electric field of over 1200 V/cm across the PD to the farthest floating diffusion (FD2 and FD4). To reduce power consumption, a TG-ON voltage of 2.4V and TG-OFF voltage of 0.4V were used while maintaining good electron sorting efficiency. Fig. 3-4 (b) shows the potential at the demodulation gates while sorting the electrons to the FDs. The diagrams verify the effectiveness of the potential ladder and spiky triangular-shaped layout, and the high-speed charge modulator can collect charges within 0.8 ns, which was calculated by electron mobility in silicon (>1000 cm^{2}V^{-1}s^{-1}).
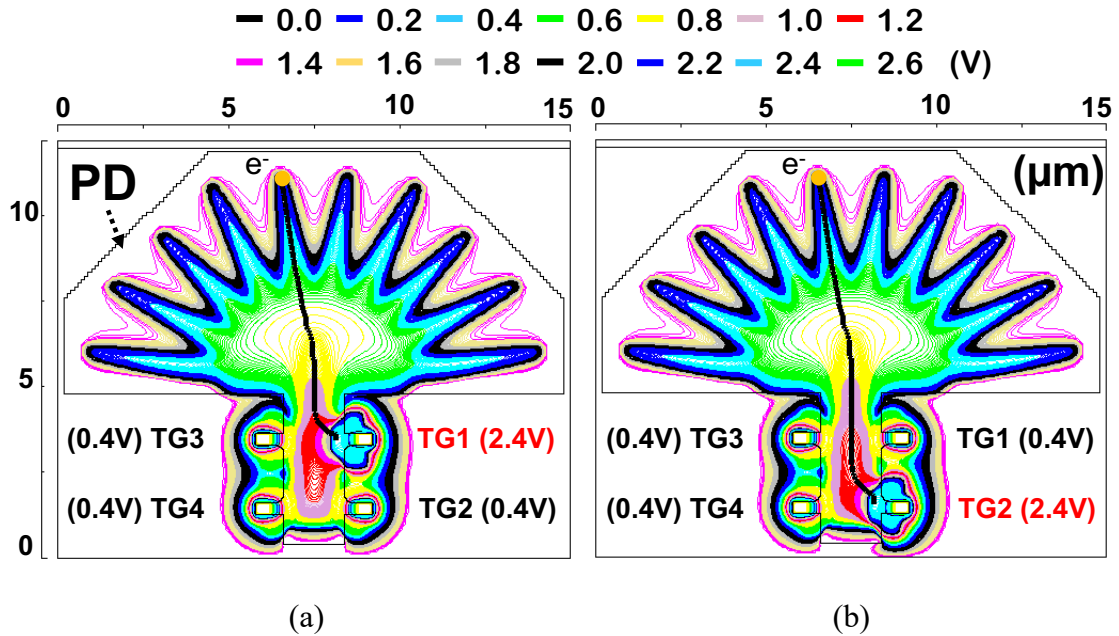
Fig. 3-3 Simulated potential diagrams with electron transfer path
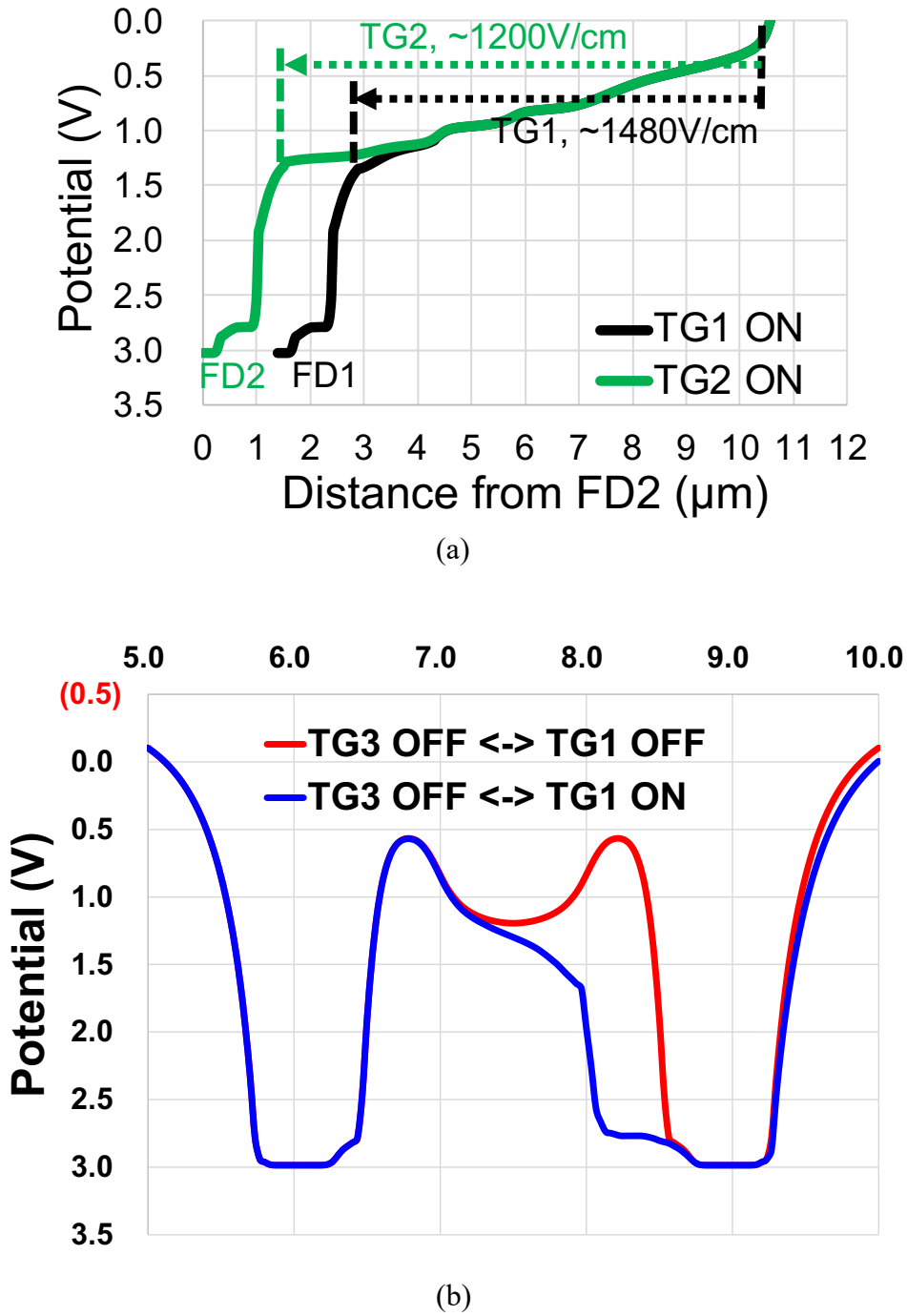**(a) TG1 turn on (b) TG2 turn on**

(a)

(b)

Fig. 3-4 Potential diagram of (a) Charge transfer path (b) TG switching

3.1.3. Sensor Operation

The timing diagram shown in Fig. 3-5 (a) illustrates the operation of the sub-frame in detail. At the beginning of each sub-frame (t1), the PD, FDs and C_{MEM} are reset using ΦTG, ΦPR, and $\Phi MR/\Phi PHs/\Phi MS$ signals, respectively. At the same time, both plates of C_{AZ} are forced to DC voltages by enabling PSF and ΦMW. The auto-zeroing operation is then performed by ΦPR and ΦPH in sequence (t2, t3) to eliminate the thermal noise and offset from FD reset. Meanwhile, the modulation is initiated with LED and TGs toggling. After the modulation is completed, the 4-tap signals are simultaneously sampled into the corresponding memories by ΦMS (t4). During the modulation, PSF and PCS are disabled 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. 3-5 (a) and (b), respectively.

➢ **High-Precision Mode (HP)**

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

➢ **High-Speed Mode (HS)**

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&PH3 and PH2&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 is performed at VMO to reduce both fixed pattern noise and low frequency noise due to MSF readout.
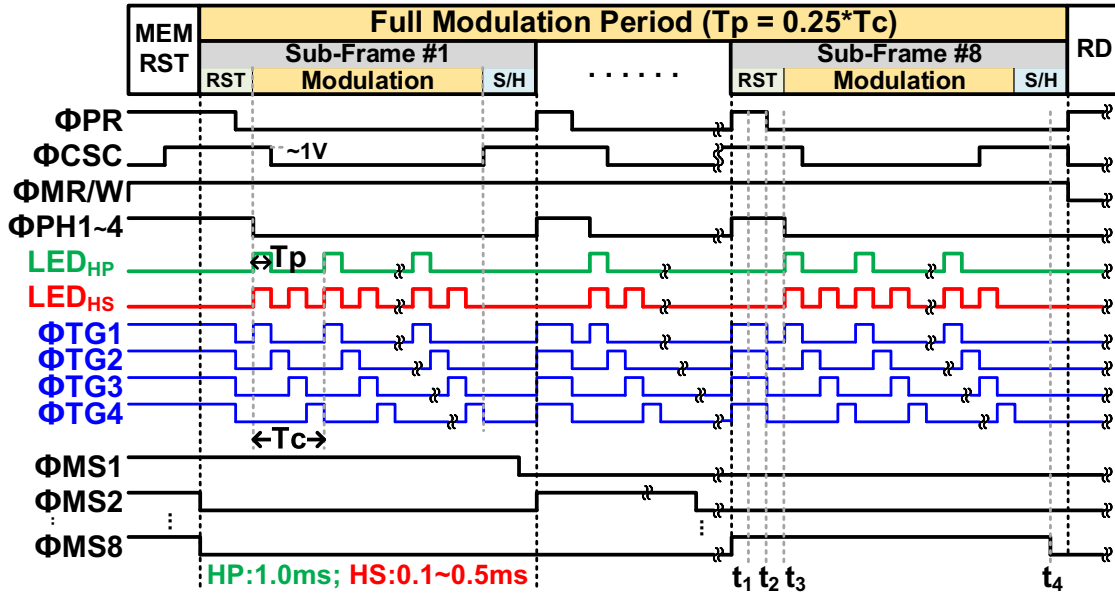
(a)
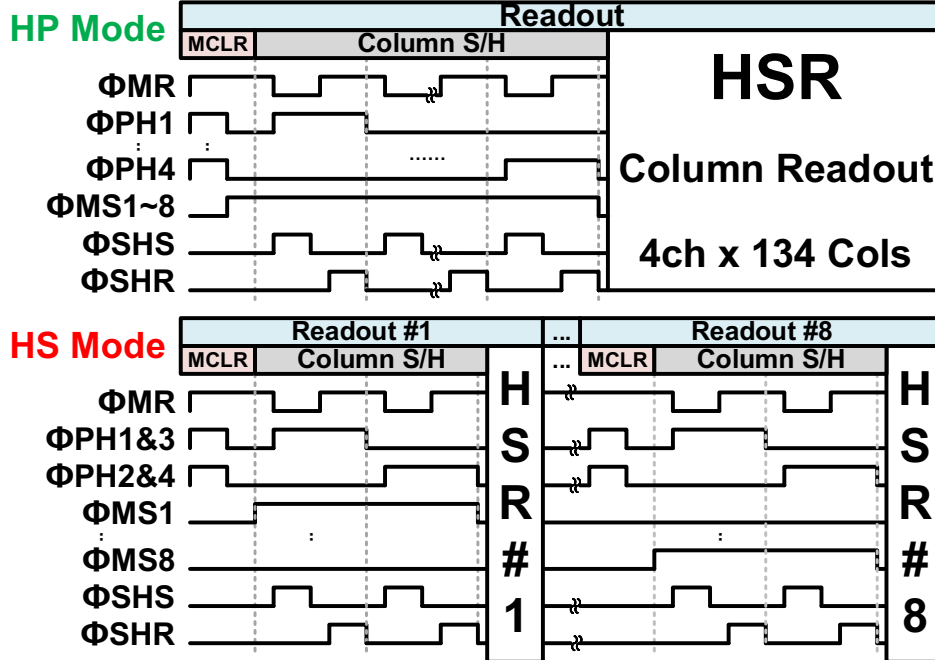
(b)

Fig. 3-5 Timing diagram of (a) Sub-frame ToF operation (b) Column readout operation of HP and HS mode.

3.2. Fabrication

The developed prototype chip was fabricated using a 0.18-µm 1-poly-Si 5-metal CIS process technology with 8-µm-thick P-epitaxial on N-substrate wafer. The power supply voltage is 3.3V/2.8V/2.4V for analog/digital/TG circuits, respectively. The chip size is 4.8 mm^H × 4.8 mm^V and the micrograph with floorplan are shown in Fig. 3-6.
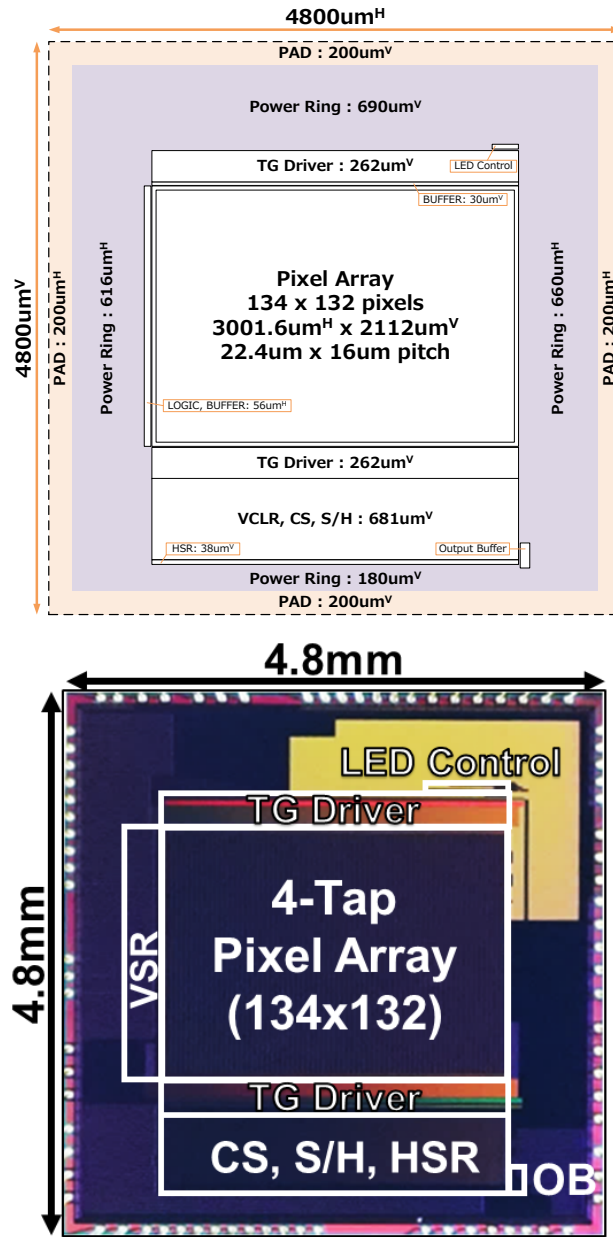

Fig. 3-6 Chip micrograph with floorplan

3.3. Verification

This section describes the verification system setup and the measurement results of the prototype iToF sensor.

3.3.1. System Noise

In this prototype sensor, a conversion gain (CG) of $84.8\ \mu V/e^-$, and a FWC around $12\ ke^-$ were confirmed. Fig. 3-7 illustrates the simplified signal readout path and its stage noise in terms of noise power. In HP mode, a FD referred noise of $4.8\ e^-$ is measured, consisting of the noise from pixel circuits ($3\ e^-$) and readout circuits ($3.8\ e^-$).
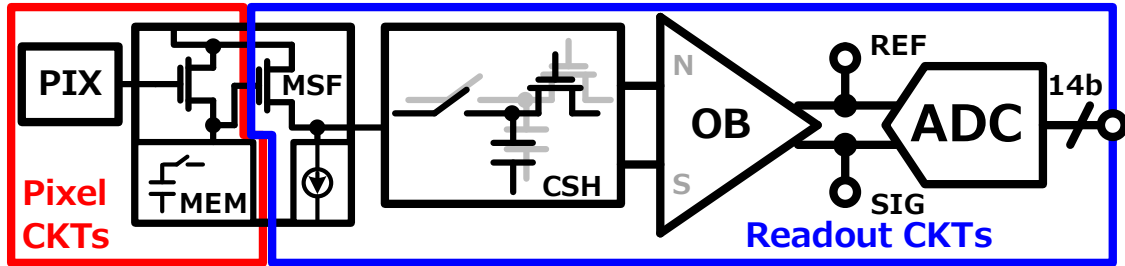
HP mode FD Refer	PIXEL	MEM	MSF	CSH	OB	ADC
Stage Noise (μV_{RMS})	215	137	131	135	154	205
Total Noise (μV_{RMS})	**407**	346	318	290	256	205

Fig. 3-7 The circuit noise of signal readout path.

The proposed HP and HS modes are evaluated and compared to conventional operation modes with high and low CG, denoted by HCG and LCG, respectively.

Fig. 3-8 compares the SNR characteristic of HCG, LCG, and HP under different illuminance levels, which can be used to indicate the maximum achievable SNR at

different distances *(d)* in an iToF system.

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 more than 8.4 dB SNR enhancement across the measured illuminance range, contributing to better depth noise performance in iToF range imaging, even at a longer distance.

Table 3-1 summarizes the basic characteristics of these modes. It should be noted that the LCG was calculated using a factor of 8, referred to HCG. Additionally, the total integrated signal electrons are proportional to the exposure time, where HP is the same as LCG and 8 times higher than HCG.

Fig. 3-8 compares the SNR characteristic of HCG, LCG, and HP under different illuminance levels, which can be used to indicate the maximum achievable SNR at different distances *(d)* in an iToF system.

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 more than 8.4 dB SNR enhancement across the measured illuminance range, contributing to better depth noise performance in iToF range imaging, even at a longer distance.

Table 3-1 Characteristic summary of different modes

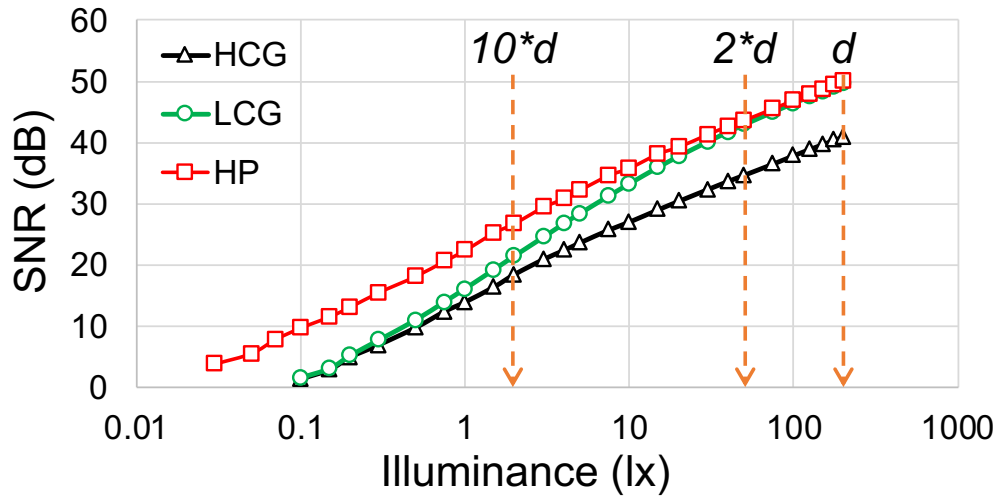|	HCG	LCG	HP	HS
CG ($\mu V/e^-$)	84.8	10.6	84.8	84.8
FWC (e^-)	12k	96k	12k	12k
Expo. (a.u.)	x1	x8	x8	x0.1 - 0.5
RN (e^-)	10.4	83.2	4.8	7.9

Fig. 3-8 SNR characteristic comparison.

The proposed HP mode exhibited a more efficient SNR enhancement capability than the LCG mode. However, please note that the prototype chip did not implement the sensor with LCG pixels. Therefore, for the following depth characterization and demonstration, the calculated LCG mode and the measured HP and HCG modes will be used.

3.3.2. Demodulation Contrast

Fig. 3-9 (a) shows the 4-tap demodulation characteristic obtained by shifting the delay between light pulse and TG1 pulse. A pulse width of 10 ns was applied in this measurement. Ideally, all signal electrons should be collected in the designated tap, but due to the limited charge transfer efficiency, some charges may remain and be collected by the following tap, leading to a degradation in the demodulation contrast (DC).

Due to the asymmetrical structure of the charge modulator, electrons being sorted to Tap2/4 experience a longer electron transfer path and weaker electric field compared to those being sorted to Tap1/3. As a result, the amplitude of the signal collected by Tap2/4 is lower than that Tap1/3.

The DC characteristic of the proposed 4-tap iToF sensor with SP modulation has been evaluated and plotted in Fig. 3-9 (b) with different demodulation pulse widths (T_P) and was calculated by Eq. 3-1 [42].

$$DC_n = \max\left(\frac{2S_n - S_{SUM}}{S_{SUM}}\right), where \ S_{SUM} = \sum_{n=1}^{4} S_n \qquad \text{Eq. 3-1}$$

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

An averaged 4-tap DC of 78.7% was obtained with a 10 ns demodulation pulse by analyzing the data from 10x10 pixels in the center of the array. However, there is a 14% difference in DC between Tap1/3 and Tap2/4 due to the unbalanced efficiency in electron transfer. Assuming that the transfer time difference of each tap remains constant, decreasing pulse width (T_P) exacerbates the difference in DC between the taps.
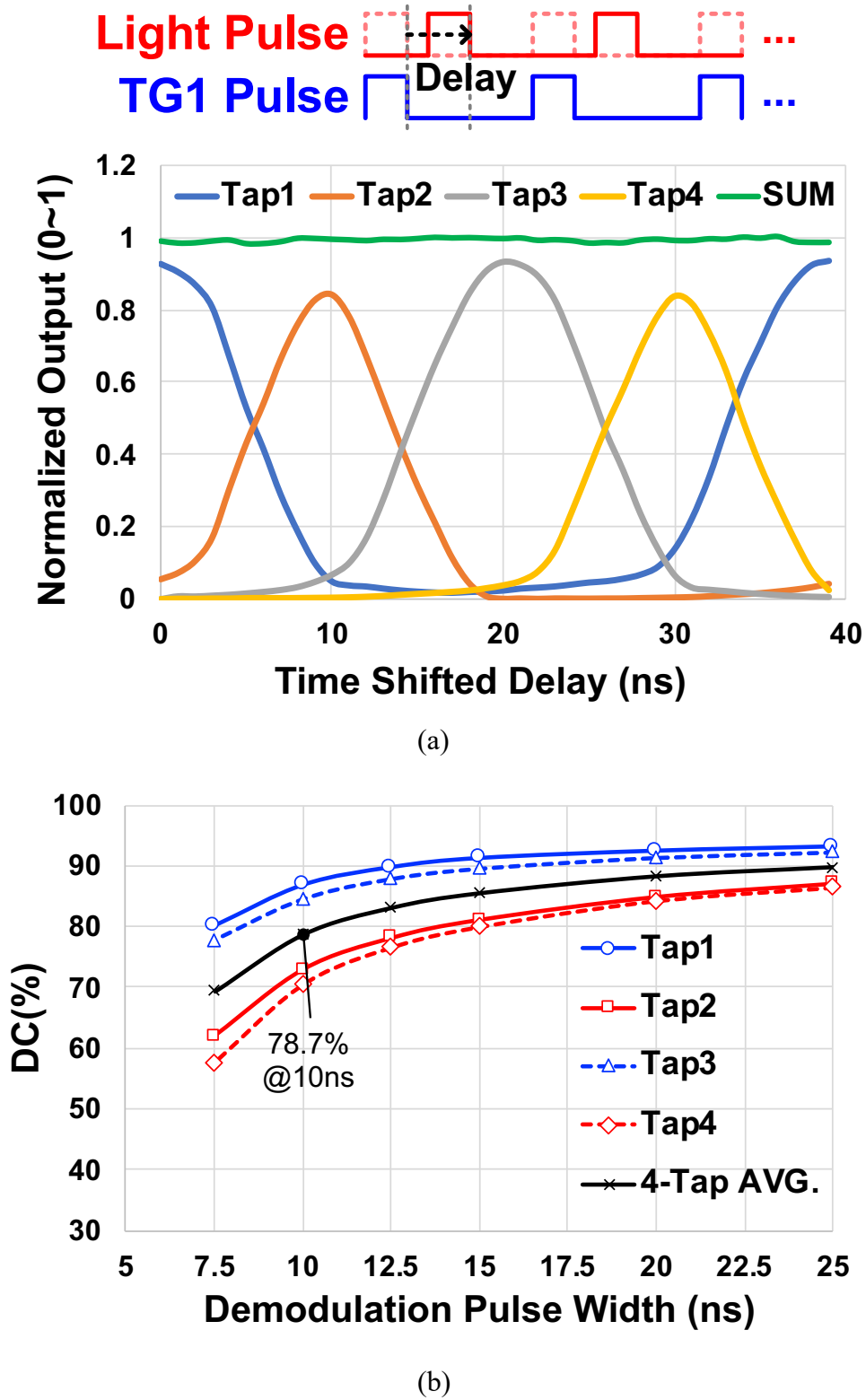
(a)

(b)

Fig. 3-9 Demodulation characteristic of tap.

(a) Delay Shifting with a Tp of 10ns (b) DC with different pulse width (Tp)

3.3.3. Measurement System

Fig. 3-10 shows the measurement setups for evaluating the depth performances. The evaluation was conducted by analyzing the central 10×10 pixels over 100 consecutive frames. The target used in the setup was a 90% reflective white flat board, and the system utilized an 850 nm vertical-cavity surface-emitting laser (VCSEL) with a peak power of 8-W to generate the modulation light pulse with a T_P of 10 ns. An IR bandpass filter was used along with an F/1.4 lens, and the measurement was carried out indoors with a fluorescent light intensity of less than 500 lx.

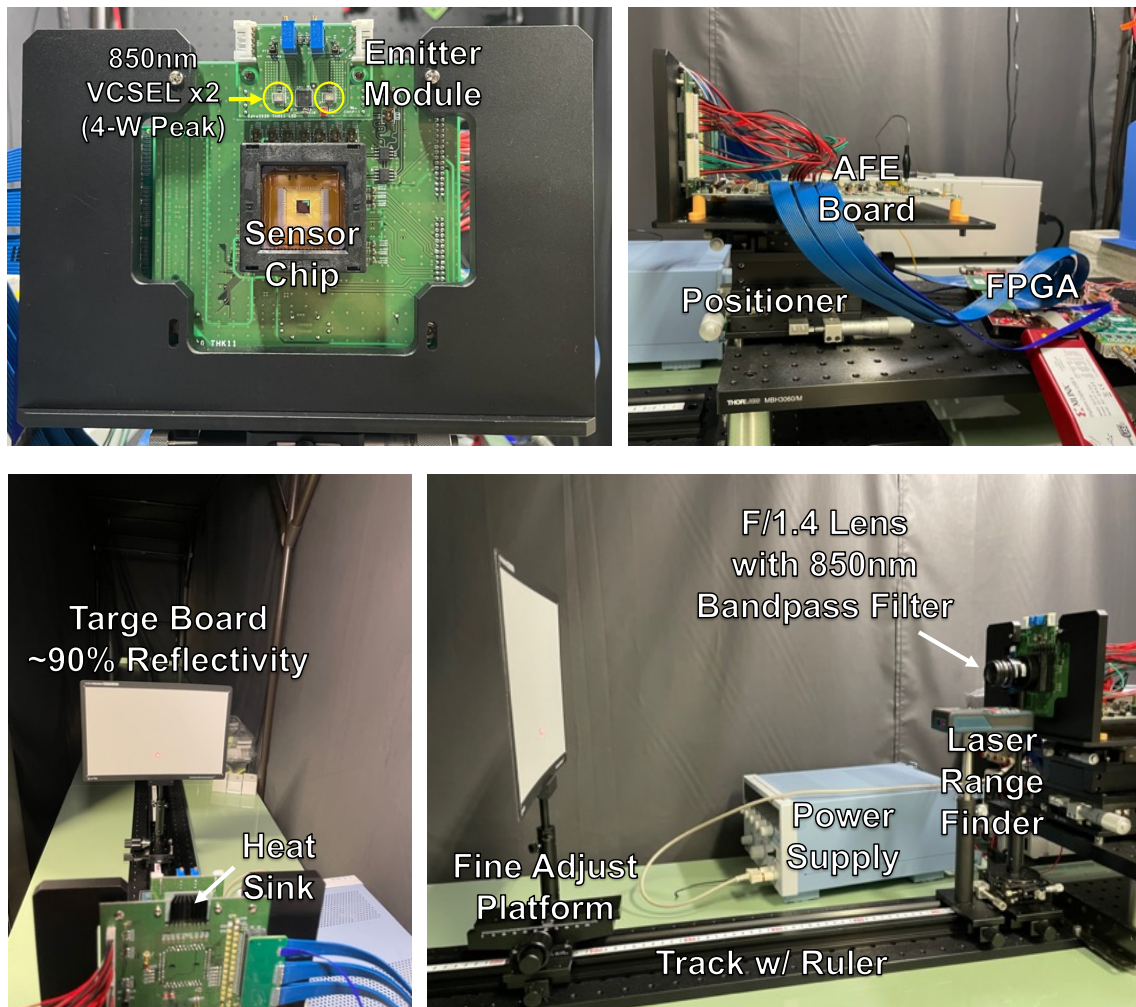The detailed system parameters are summarized in Table 3-2.

Fig. 3-10 Measurement setups for depth performance characterization.

Table 3-2 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]
Emitter	Emitter type	VCSEL with a FoI of 60° × 45°
	Wavelength	850 nm with a bandwidth of 2 nm
	Optical power	8-W peak power (2 × 4-W) 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 time-window (TW). The HP operation offers 4-TW with a cycle time (T_C) of 40 ns, whereas the HS mode has only 1-TW due to the pseudo-2-tap operation. Note that there was a time delay of approximately 1 ns between the TG1 pulse and the light pulse. The slope and time offset of the measured depth were calibrated for each TW.

3.3.4. Depth Accuracy and Precision

➢ **High Precision Mode**

The theoretical depth noise of the proposed HP mode was compared. The equation can be expressed by Eq. 3-2,

$$\sigma_{HP} = \frac{c}{8f_m} \frac{\sqrt{\dfrac{N_\varphi(1-R_S)}{N_{SubF}} + 2(\dfrac{N_B}{N_{SubF}} + RN_{HP}{}^2)(1 - 3R_S + 3R_S^2)}}{DC \cdot N_S} \qquad \text{Eq. 3-2}$$

where the measured readout noise is denoted by RN_{HP}. Here, f_m of 25 MHz, N_B of 0, RN_{HP} of 4.8, N_{SubF} 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 8,000 in Tap2.

Fig. 3-11 (a) and (b) show 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-fps 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.
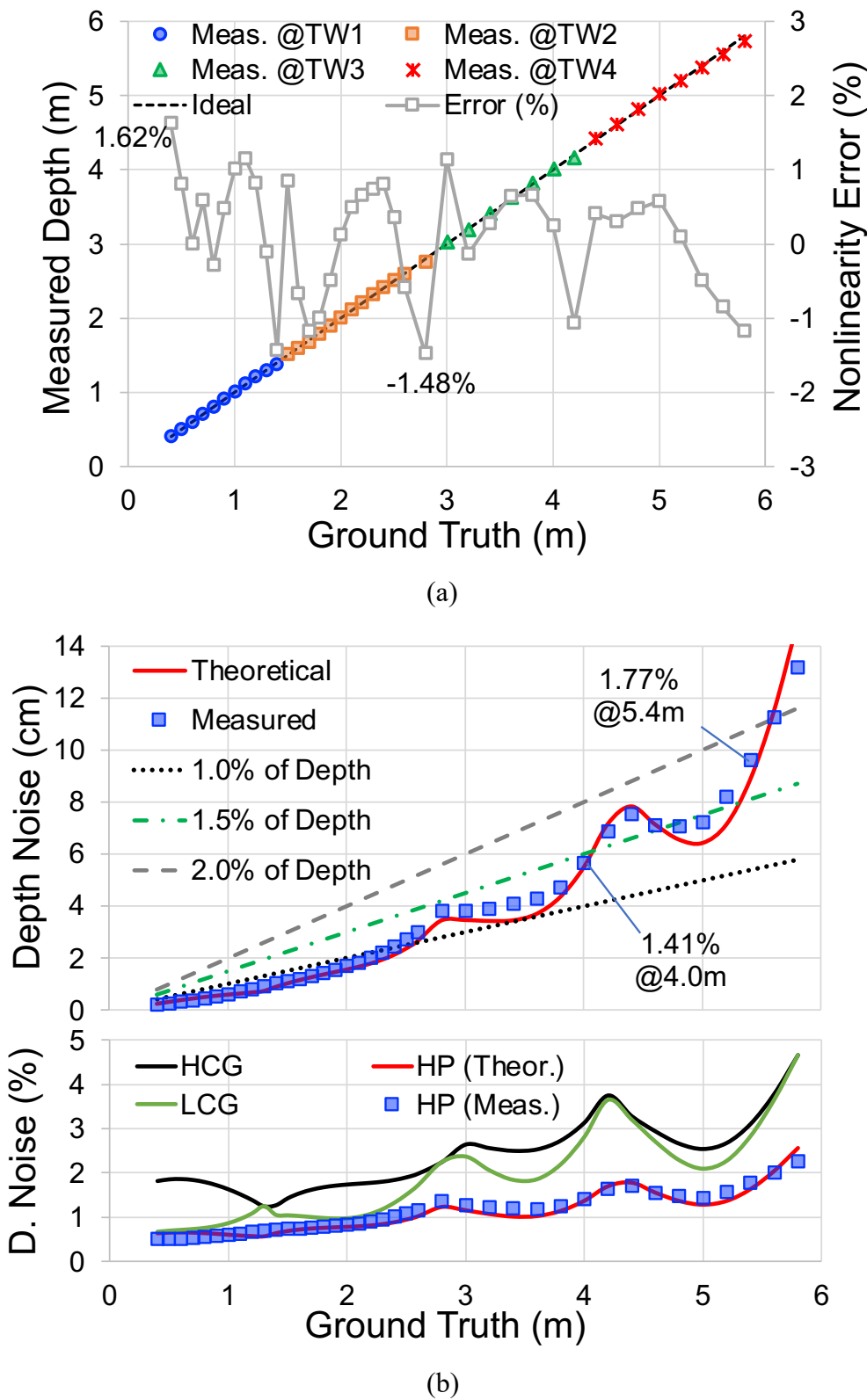
(a)

(b)

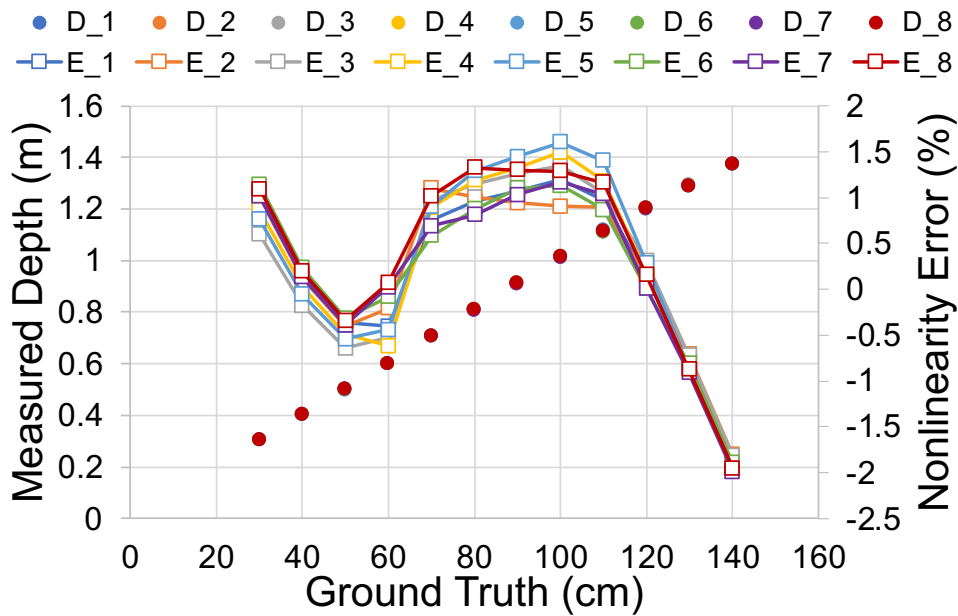**Fig. 3-11 (a) Depth accuracy (b) Depth precision of HP mode**

➢ **High Speed Mode**

The equation of HS mode, which was calculated with pseudo-2-tap operation, can be expressed by Eq. 3-3,

$$\sigma_{HP} = \frac{c}{8f_m} \sqrt{\frac{\frac{N_\varphi(1-R_S)}{N_{SubF}} + 2(\frac{N_B}{N_{SubF}} + RN_{HP}{}^2)(1-3R_S+3R_S^2)}{DC \cdot N_S}} \qquad \text{Eq. 3-3}$$

where the measured readout noise is denoted by RN_{HS}. Here, f_m of 50 MHz, N_S of 16400, N_B of 0, RN_{HS} of 7.9, DC of 0.72 were applied.

The depth performance of each subframe was characterized separately. Fig. 3-12 (a) and (b) show 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 8-subframe averaged depth nonlinearity was <1.91%, with depth noise <1.82 for the range of 0.3-1.4 m.
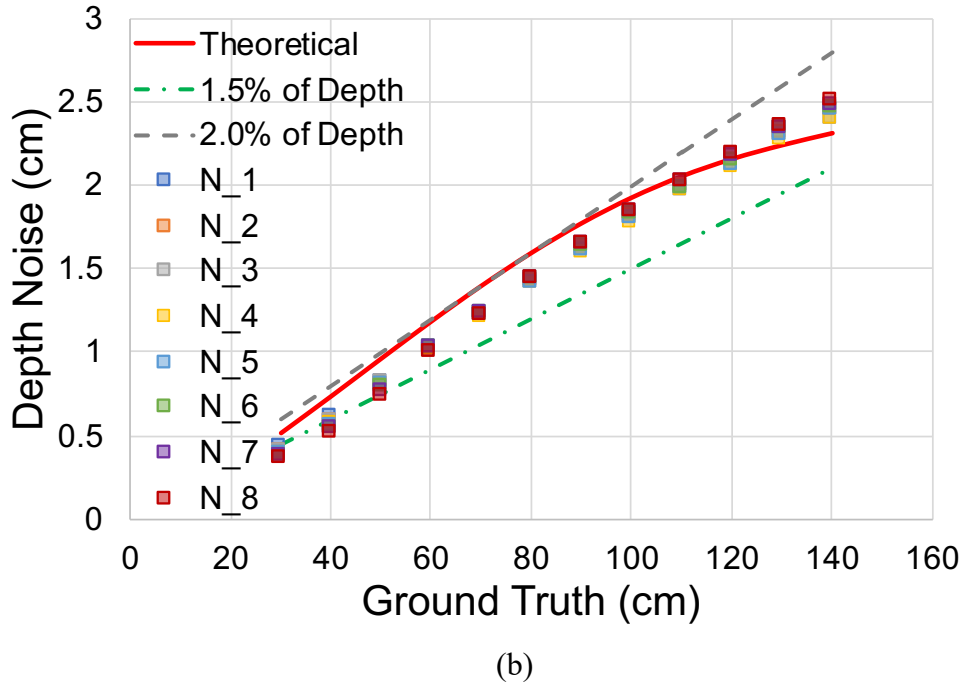

(a)

(b)

Fig. 3-12 (a) Depth accuracy (b) Depth precision of HS mode

The noise curves for both HP and HS modes, calculated based on measured system parameters, provide a reliable estimate of expected depth noise. However, it is important to note that in practical iToF measurements, several limitations and uncertainties can affect depth performance, such as clock distortion and jitter, limited bandwidth of modulated light pulse for SP modulation, unstable sensor and emitter power, and heat buildup during sensor operation. To achieve relatively stable measurement results, this prototype sensor used external voltage sources and an ADC. Additionally, a heat sink was attached to the designed emitter module to address heat buildup during operation.

The figure-of-merit (FoM) can be calculated using Eq. 3-4 [11], where the depth noise is the maximum value measured in the selected range in terms of percentage (%), and the pixel rate (PR) is calculated using Eq. 3-5. In HS mode, the equivalent pixel rate is expressed as Eq. 3-6.

$$FoM = \frac{power \times depth\ noise}{pixel\ rate(PR) \times N_{Tap}} \left[\frac{pJ}{pixel}\right] \qquad \text{Eq. 3-4}$$

$$RR = frame\ rate \times pixel\ count \left[\frac{pixel}{s}\right] \qquad \text{Eq. 3-5}$$

$$PR_{HS} = \frac{N_{SubF} \times pixel\ count}{refresh\ period} \left[\frac{pixel}{s}\right] \qquad \text{Eq. 3-6}$$

The conventional FoM offers a quantitative way to measure the performance in terms of power efficiency and ranging precision. However, as depicted in Fig. 3-13, the efficiency diminishes as the distance increases due to the growing depth noise. Therefore, it is necessary to include the distance information in the FoM equation to identify the optimal working range that yields the best efficiency for an iToF imager.
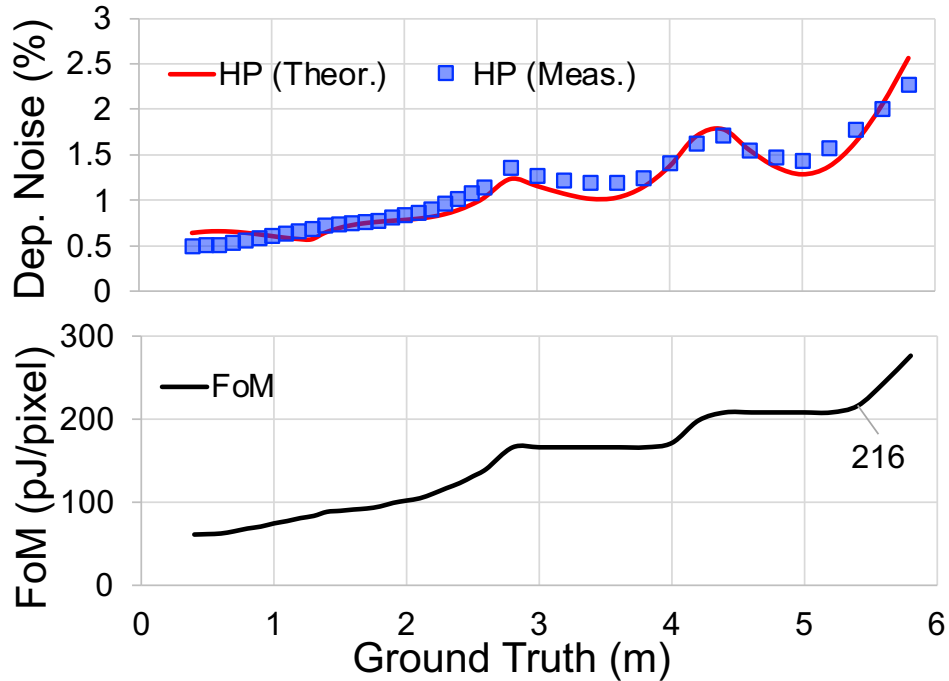

Fig. 3-13 Conventional FoM over distance.

In an iToF ranging system, the modulation period is adjusted based on the signal saturation level at the minimum distance. Meanwhile, the power of the modulated light decreases rapidly with depth, limiting the maximum distance due to low SNR. Hence, the detection range ratio (DDR) is incorporate into the range-FoM (R-FoM) equation which is defined as Eq. 3-7 and plot in Fig. 3-14.

$$R\text{-}FoM = \frac{FoM}{Detection\ range\ ratio\ (DRR)} \left[\frac{pJ}{pixel}\right] \qquad \text{Eq. 3-7}$$

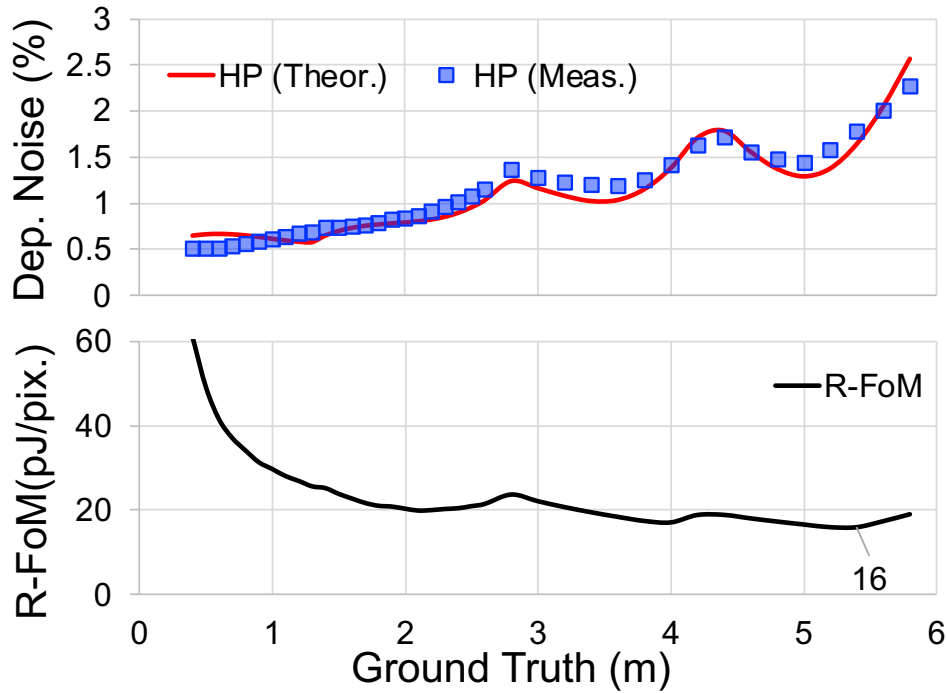$$DRR = (Distance_{MAX}/Distance_{MIN}) \qquad \text{Eq. 3-8}$$

Fig. 3-14 Range FoM over distance.

By adopting the R-FoM, the optimal working range with the best efficiency for the proposed iToF imager is 16pJ/pixel at the range of 0.4-5.4 m.

Table 3-3 summarizes the performance of the developed iToF sensor.

Table 3-3 Performance of the developed iToF range imager.

Process Tech.	0.18μm FSI	
Pixel pitch(μm)	22.4×16	
Pixel archit.	4-tap	
Pixel array	134×132 [1]	
Modulation speed	25MHz (HP), 50MHz (HS) (10ns short-pulse)	
Demodulation contrast	78.7% @ 10ns PW [1]	
Conv. Gain	85μV/e-	
Frame rate(fps)	90 (HP)	2K-10K (HS)
Read noise	407μV	670μV
Depth noise	<1.77% @ 0.4~5.4m	<1.82% @ 0.3-1.4m [2]
Chip power	77.8mW	40.1mW [2]
FoM(pJ/pixel)	216	86
R-FoM(pJ/pixel)	16	18.4
[1] 4-tap averaged DC; [2] Measured at 2 Kfps with 33.3ms refresh period;		

3.3.5. Sample Images

The experimental setup for demonstrating the high-precision mode (HP) is shown in Fig. 3-15 (a). The sensor was operated at 90 fps in an indoor environment with < 500-lx fluorescent light. The nearest object was placed at a distance of 45 cm, and an alphabet "U" was positioned 15 cm in front of the background panel, at a distance of 2.55 m.
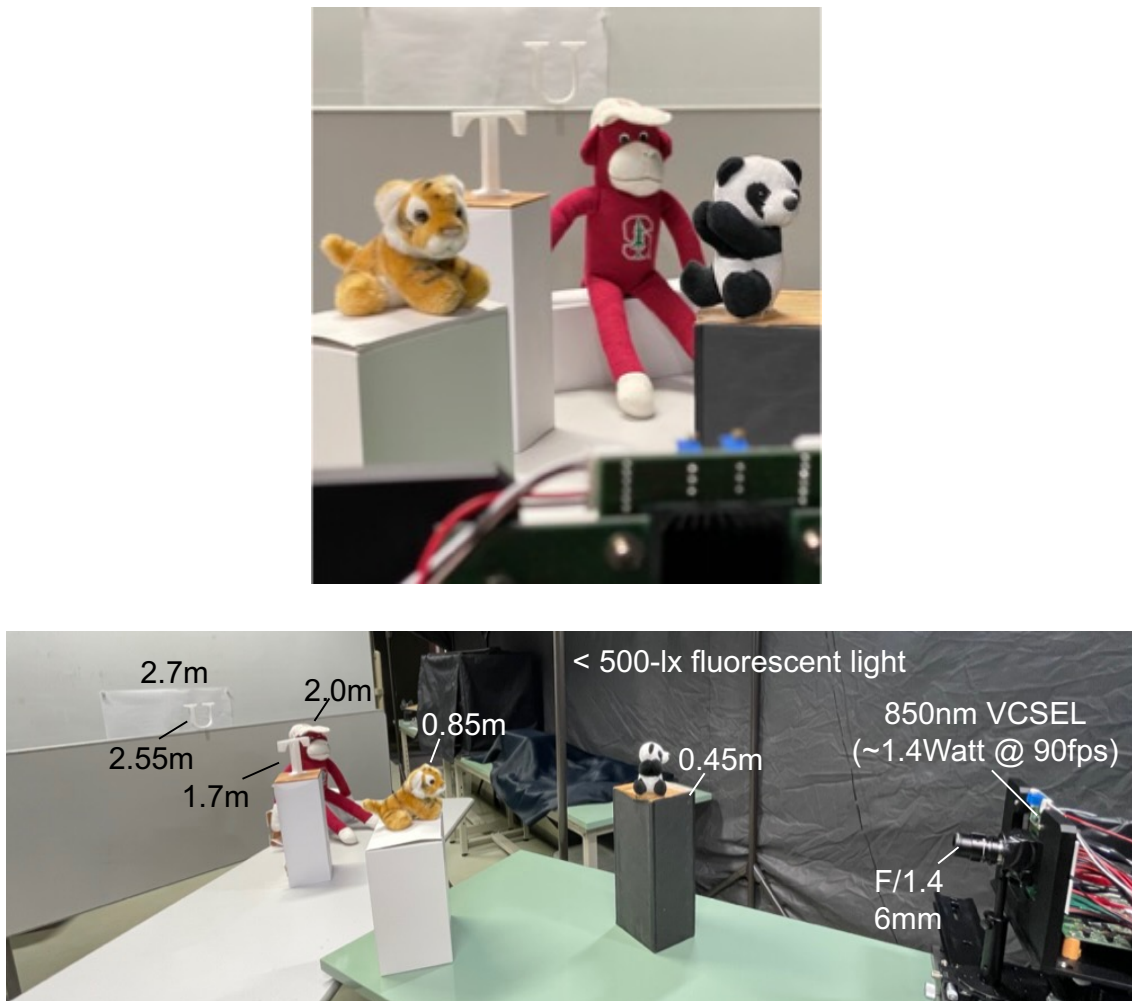

Fig. 3-15 Experimental setup for HP mode demonstration.

Fig. 3-16 shows the captured images of HP mode. The intensity map demonstrated a rapid decrease in reflected light as the distance increased, leading to poor depth resolution at longer ranges when using HCG without subframe ToF operation.

In contrast, 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 [43].
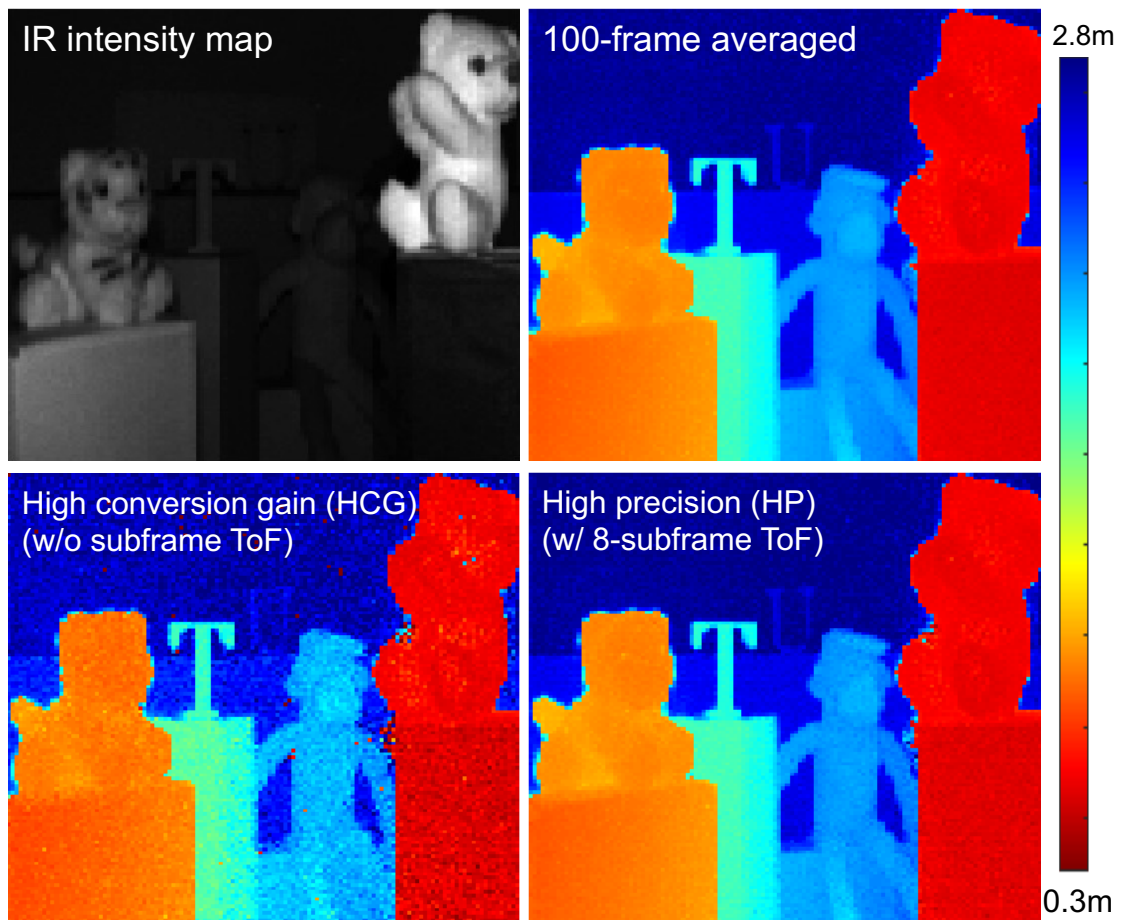

Fig. 3-16 Captured images of HP mode at 90 fps.

On the other hand, the scenes captured using the high-speed mode (HS) are shown in Fig. 3-17. In the first scenario, an alphabet "T" was spinning at a speed of around 4500rpm, obstructing the objects behind it. In the second scenario, a bouncing plate propelled by rubber bands was used to create rapid movement towards the sensor. Without the HS mode readout operation, the rapid motions in both scenes induced serious motion blur, making it difficult to recognize the pattern and direction of the moving objects.

(a)

(b)

(c)

(d)
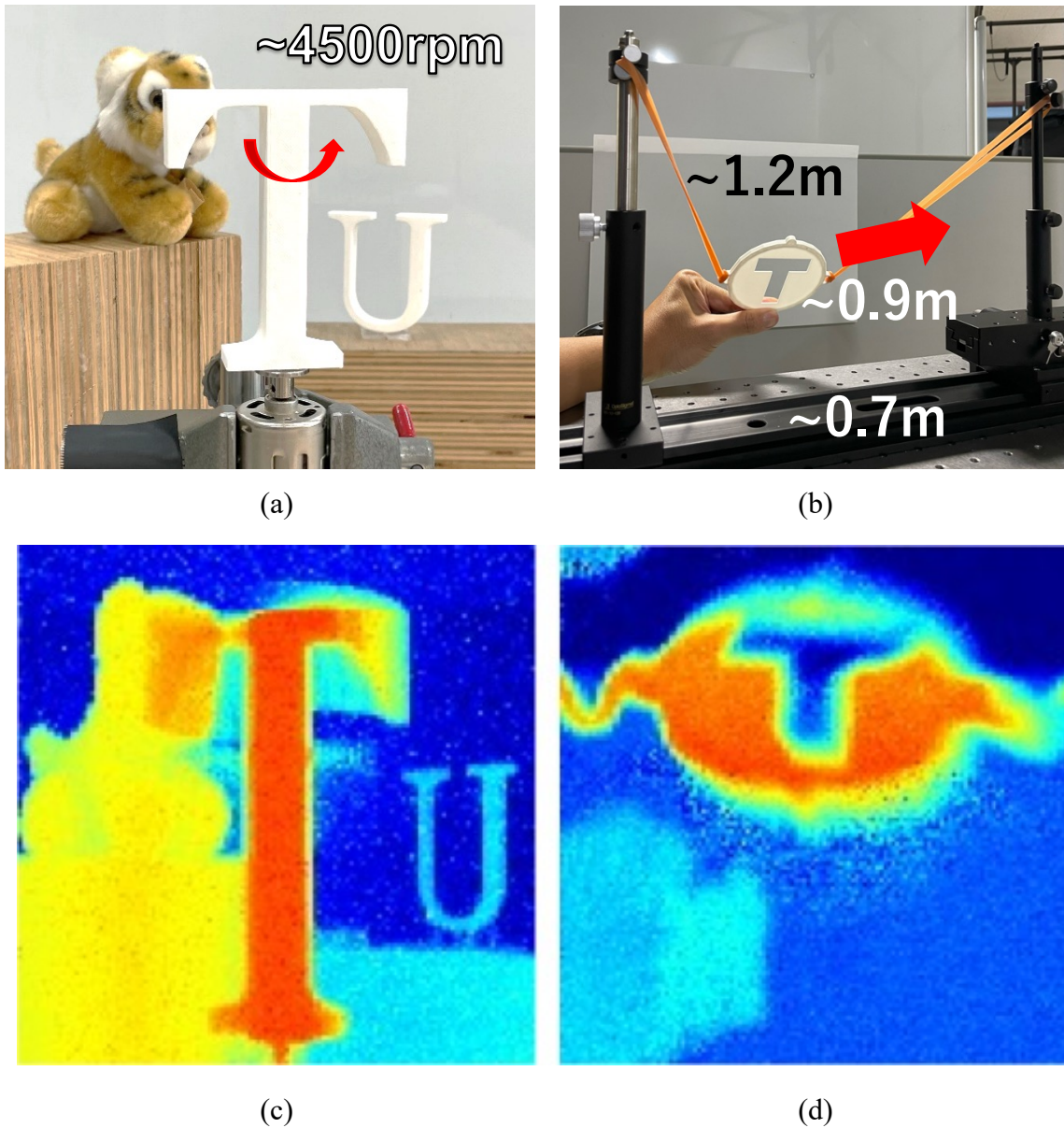
Fig. 3-17 Experimental setup for HS mode demonstration.

The HS mode allows the captured frame to be disassembled into 8 burst frames, as shown in Fig. 3-18. These images were captured at 2 Kfps and 4 Kfps with 0.5-ms and 0.25-ms temporal resolution, respectively. The HS mode provides clear depth images of the background objects and observing target during high-speed motion, enabling accurate analysis and recognition.
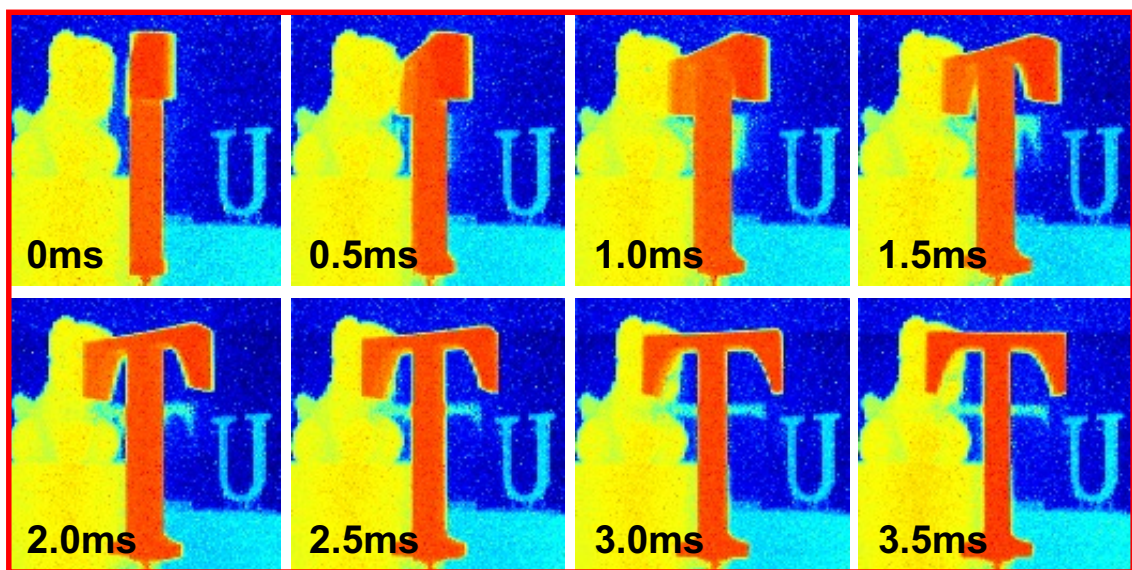
(a)
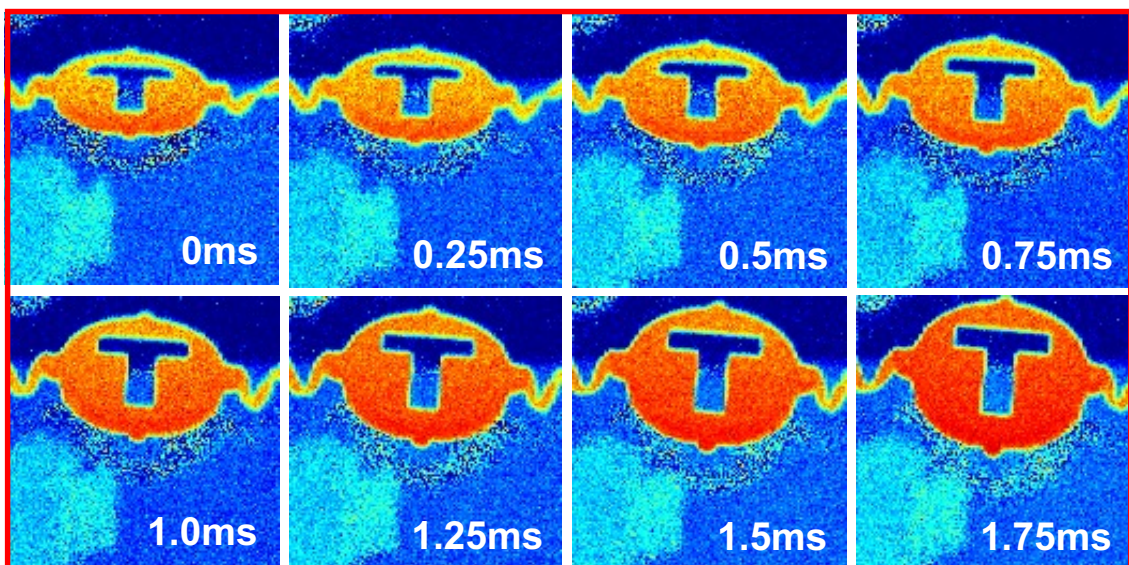
(b)

Fig. 3-18 Captured images of HS mode at (a) 2 Kfps (b) 4 Kfps

3.4. Applicability of the Proposed Range Imager

The development of iToF sensors in both HP and HS modes offers a range of benefits for various applications. In practical settings, the HP mode is well-suited for constructing 3D environments for human behavior analysis. The features of low depth noise, suppressed motion artifact, and wide-range imaging capability enable precise and accurate measurements, making it useful in applications such as gesture control, VR/AR, and security monitoring system. The HS mode, on the other hand, provides high temporal resolution images that are suitable for machine vision tasks such as rapid motion recognition and analysis.

Moreover, the combination of HP and HS modes offers significant potential in event-driven applications such as comprehensive automotive safety systems. For example, as shown in Fig. 3-19, 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.
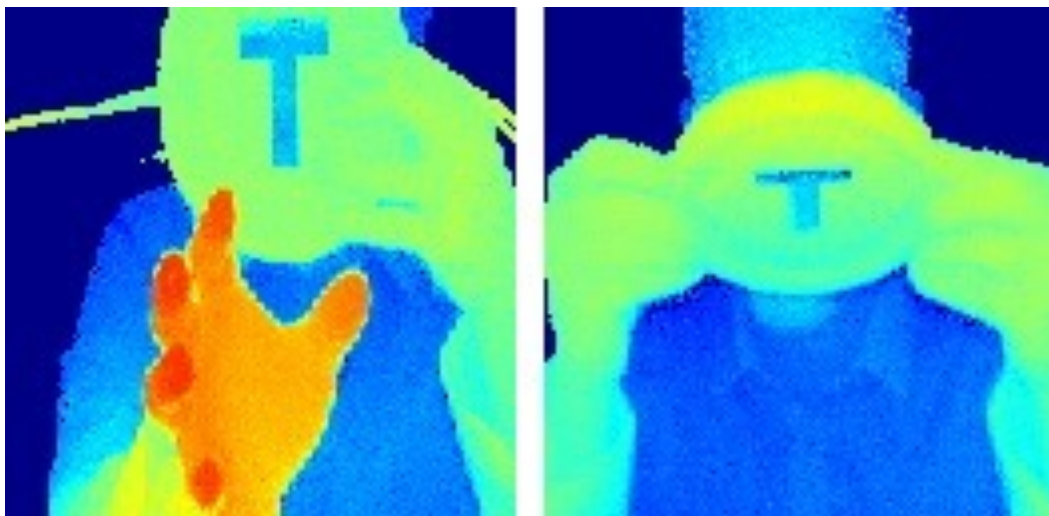

Fig. 3-19 HP mode for behavior recognition.

In the event of severe accidents, where the deceleration forces can exceed 50-G and cause the human body to come to a sudden halt within a couple milliseconds. At this moment, a high-speed burst imaging is needed to record the crucial spatial information.

The experiment shown in Fig. 3-20 (a) and (b) indicate the successful application of HS mode imaging in capturing the roll-over and impact accidents of a car at 5 Kfps and 10 Kfps, respectively. The imaging method allowed for the observation of the pattern on the target as well as the rotation and displacement movement of the target. Besides, by analyzing the movement among the recorded frames, the velocity of the target can be calculated. These information can be used by an automotive safety system to analyze depth images and detect any abnormal events, triggering emergency support if necessary.

The captured burst images can then be used to analyze the rapid motion and detect any potential injuries that might have occurred. Additionally, 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.
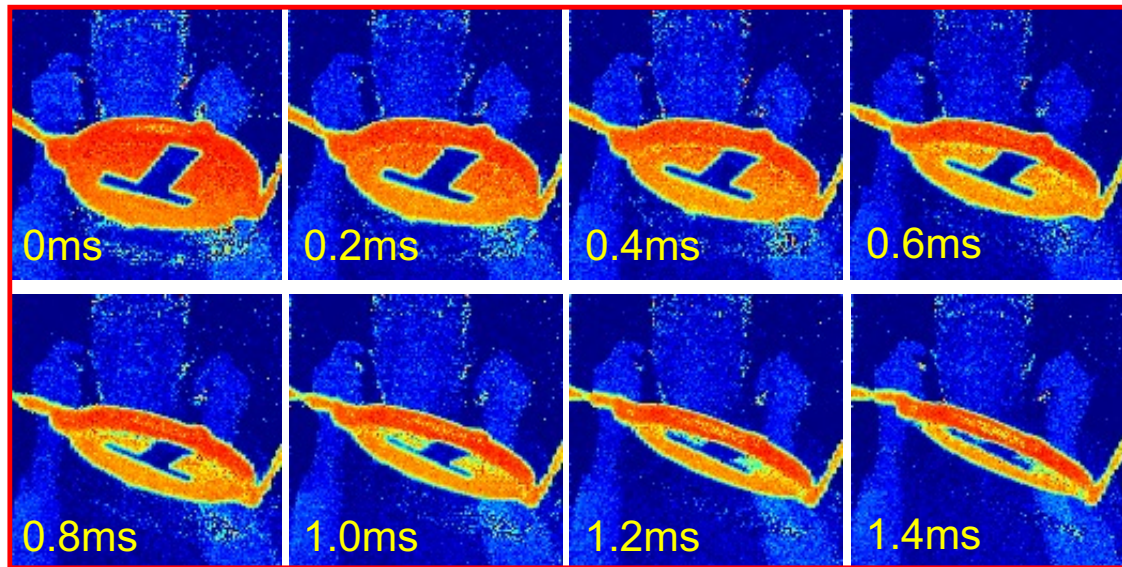
(a)
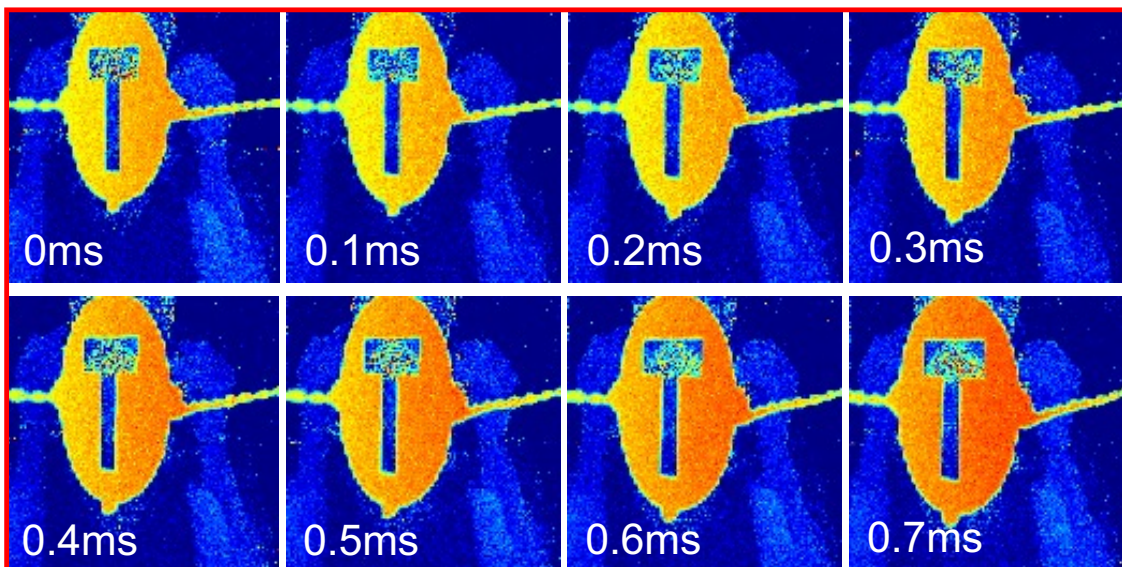
(b)

Fig. 3-20 Demonstration of a scenario for the car accident using the proposed sensor with HS mode at (a) 5 Kfps and (b) 10 Kfps.

3.5. Discussions

3.5.1. Sensor Improvement

Firstly, the sensor resolution can be significantly enhanced by introducing backside illumination (BSI) and 3D stacking technologies, which would improve the layout flexibility and area efficiency to achieve higher fill-factor.

Secondly, IR-responsivity enhancing technologies such as micro-lenses, deep trench isolation (DTI) [44], [45], and pyramid surfaces for diffraction (PSD) structures [46], [47] can be adopted to reduce the exposure time, resulting in higher frame speed and lower power consumption. Thirdly, to achieve better DC, creating a doping gradient on p-epi would be beneficial in increasing the charge collection efficiency in the vertical direction, resulting in better depth precision.

Lastly, the readout noise can be reduced by implementing a column-ADC circuit and using higher density in-pixel memory with textured deep trench SiN capacitors [48] or high-capacity DRAM capacitors [49], [50]. Moreover, by utilizing high-density capacitors and 3D stacking techniques, it is possible to implement more in-pixel memory, which can extend the duration of burst imaging even at higher frame rates. Note that the frame time of HS mode is only limited by memory reset 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 rates, and better ranging performance for the development of high-speed and high precision iToF depth imagers.

3.5.2. Automotive Application

The high precision and high framerate of the developed 3D imager make it a promising technology for various automotive applications, such as behavior monitoring and event-driven accident detection. However, to be successfully applied in real-world automotive scenarios, there are several other challenges that need to be addressed:

➢ Scene adaptability: Automotive environments can vary significantly, from bright daylight to low-light conditions, and from urban streets to rural roads. The 3D imager must demonstrate robust performance across different lighting conditions and adapt to various types of scenes to ensure accurate depth sensing and object detection.

➢ Signal interference: In automotive settings, there may be various sources of light interference, such as streetlights, reflections from shiny surfaces, and especially from ranging cameras of other vehicles. The 3D imager should be able to effectively filter out unwanted light sources and focus on the relevant depth information to avoid inaccuracies in the measurements.

➢ Temperature robustness: Vehicles are exposed to a wide range of temperatures, from extremely cold to hot climates. The 3D imager must be designed for temperature variations and maintain its performance under challenging thermal conditions.

➢ Fast data processing: Automotive applications demand real-time processing capabilities, especially for safety-critical tasks like accident detection and collision avoidance. The 3D imager should provide fast and efficient data processing to enable quick response times and decision-making.

Addressing these challenges can provide the way for safer and more advanced automotive applications in the future.

3.6. Summary of Chapter. 3

This chapter described the details of the implementation and verification of the prototype iToF sensor with high-precision and high-speed depth imaging capability.

The pixel array consists of $134^H \times 132^V$ pixels, with a pixel size of 22.4^H µm$\times 16^V$ µm. Each pixel has a 4-tap high-speed charge modulator and a 4×8 memory array. The sub-frame ToF operation is constructed by 8 subframes with SP modulation. High-precision mode (HP) imaging is performed by giving 4-tap modulation with memory averaging readout. In contrast, high-speed mode (HS) performs pseudo-2-tap operation, then obtains the burst images by readout the memory individually.

The proposed iToF imager achieved high-performance results, including up to 10 Kfps range imaging with the HS mode and less than 1.77% depth noise for a range of 0.4-5.4 m with the HP mode. Additionally, the best efficient working range was determined by range-FoM (R-FoM), achieving 16pJ/pixel with the range of 0.4-5.4 m.

The potential application of the iToF imager in automotive safety systems was demonstrated, as it can be used for behavior monitoring and accident detection. This development represents a new direction for high-speed 3-D imaging applications in machine vision and beyond.

Chapter. 4.

Half Pulse 2-Tap 4-Phase iToF Ranging Method with Sub-Frame Operation

4.1. Conventional 2T-4PH iToF Sensor

The 2-tap pixel architecture has been a popular choice to realize an iToF ranging system due to the structure simplicity and better matching between taps. However, due to the requirement of background light cancelling (BGLC), the 2-tap iToF needs two modulation periods to acquire 4-phase signals.

where c is the speed of light, and R_{SP} is defined as time-shift ratio. For the conventional SP modulation, R_{SP} can be obtained by Eq. 4-3 and Eq. 4-4.

$$R_{SP} = \begin{cases} \dfrac{1}{2}(1 + \dfrac{\Delta Q_0}{|\Delta Q_0| + |\Delta Q_{\pi/2}|}) \,, if \, \Delta Q_{\pi/2} \geq 0 \\ \dfrac{1}{2}(3 + \dfrac{-\Delta Q_0}{|\Delta Q_0| + |\Delta Q_{\pi/2}|}) \,, if \, \Delta Q_{\pi/2} < 0 \end{cases} \qquad \text{Eq. 4-3}$$

$$\begin{aligned} \Delta Q_0 &= Q2_0 - Q1_0 \\ \Delta Q_{\pi/2} &= Q2_{\pi/2} - Q1_{\pi/2} \end{aligned} \qquad \text{Eq. 4-4}$$

where the differential charges, ΔQ_0 and $\Delta Q_{\pi/2}$, are used to cancel the offset from circuit and system, regardless of the BGL level, ensuring a reliable iToF depth calculation.

Fig. 4-1 shows the operation principle of conventional 2-tap 4-phase (2T-4PH) operation with continuous square pulse (SP) modulation[28], [30]. The modulation cycle time and pulse width are denoted by T_C and T_{SP}, respectively, where $T_{SP} = T_C/2$. The 4-phase operation are achieved by two modulation periods, $PH(0, \pi/2)$, with $\pi/2$ phase shift. The 2-tap sampling clocks and the integrated charge signals are represented by $Tap1(0, \pi/2)$, $Tap2(0, \pi/2)$, $Q1(0, \pi/2)$ and $Q2(0, \pi/2)$, respectively.

The distance (d) between the sensor and object can be obtained by the light traveling time (T_{ToF}),

$$Distance(d) = \frac{c}{2} \times T_{ToF} \qquad \text{Eq. 4-1}$$

$$T_{ToF} = T_{SP} \times R_{SP}, where \ 0 \leq R_{SP} < 2 \qquad \text{Eq. 4-2}$$

where c is the speed of light, and R_{SP} is defined as time-shift ratio. For the conventional SP modulation, R_{SP} can be obtained by Eq. 4-3 and Eq. 4-4.

$$R_{SP} = \begin{cases} \frac{1}{2}\left(1 + \frac{\Delta Q_0}{|\Delta Q_0| + |\Delta Q_{\pi/2}|}\right), if \ \Delta Q_{\pi/2} \geq 0 \\ \frac{1}{2}\left(3 + \frac{-\Delta Q_0}{|\Delta Q_0| + |\Delta Q_{\pi/2}|}\right), if \ \Delta Q_{\pi/2} < 0 \end{cases} \qquad \text{Eq. 4-3}$$

$$\begin{aligned} \Delta Q_0 &= Q2_0 - Q1_0 \\ \Delta Q_{\pi/2} &= Q2_{\pi/2} - Q1_{\pi/2} \end{aligned} \qquad \text{Eq. 4-4}$$

where the differential charges, ΔQ_0 and $\Delta Q_{\pi/2}$, are used to cancel the offset from circuit and system, regardless of the BGL level, ensuring a reliable iToF depth calculation.
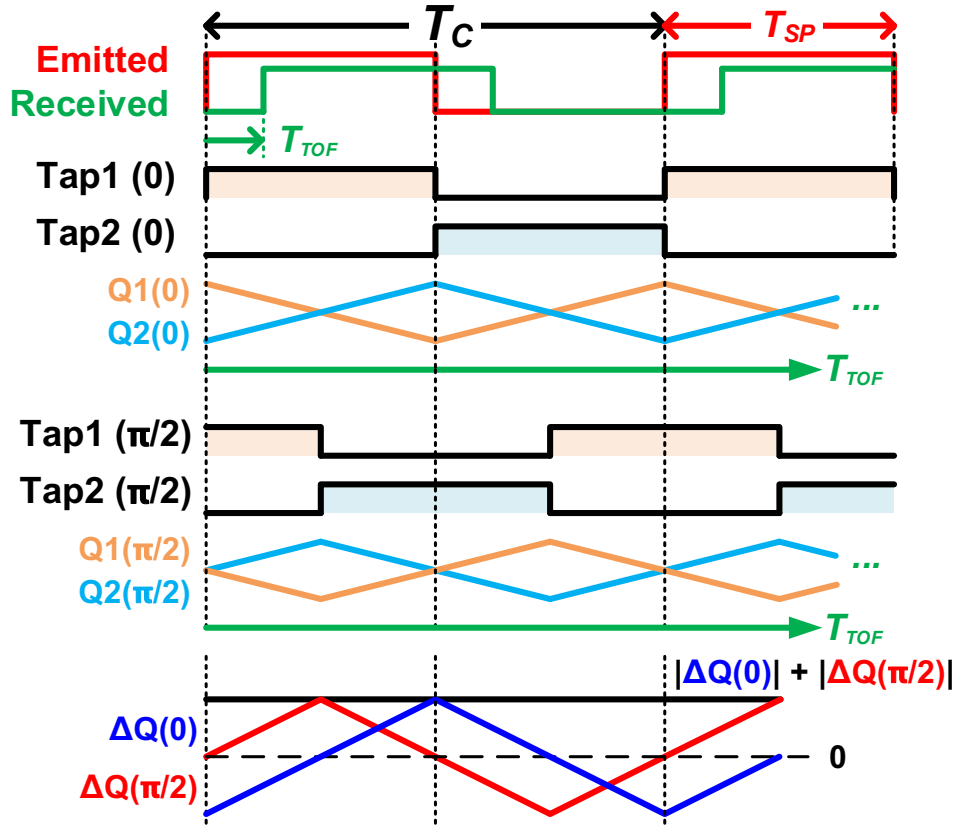
Fig. 4-1 Operation principle of conventional 2T-4PH continuous SP modulation.

By using the conventional 2T-4PH operation, the BGLC ranging result can be obtained. However, increasing the modulation frequency to improve the depth precision can result in a reduction in the unambiguous range due to phase wrapping (as discussed in Section. 1.4.3). In addition, the intensive power consumption and dissipation during the exposure period, mainly caused by the toggling demodulation gates and modulated light source, may affect the robustness and durability of the iToF system.

On the other hand, as shown in Fig. 4-2, conventional 2T-4PH operation requires temporal multiplexing by two consecutive frames to construct depth image. This frame-to-frame processing latency can result in motion artifacts when capturing moving objects.
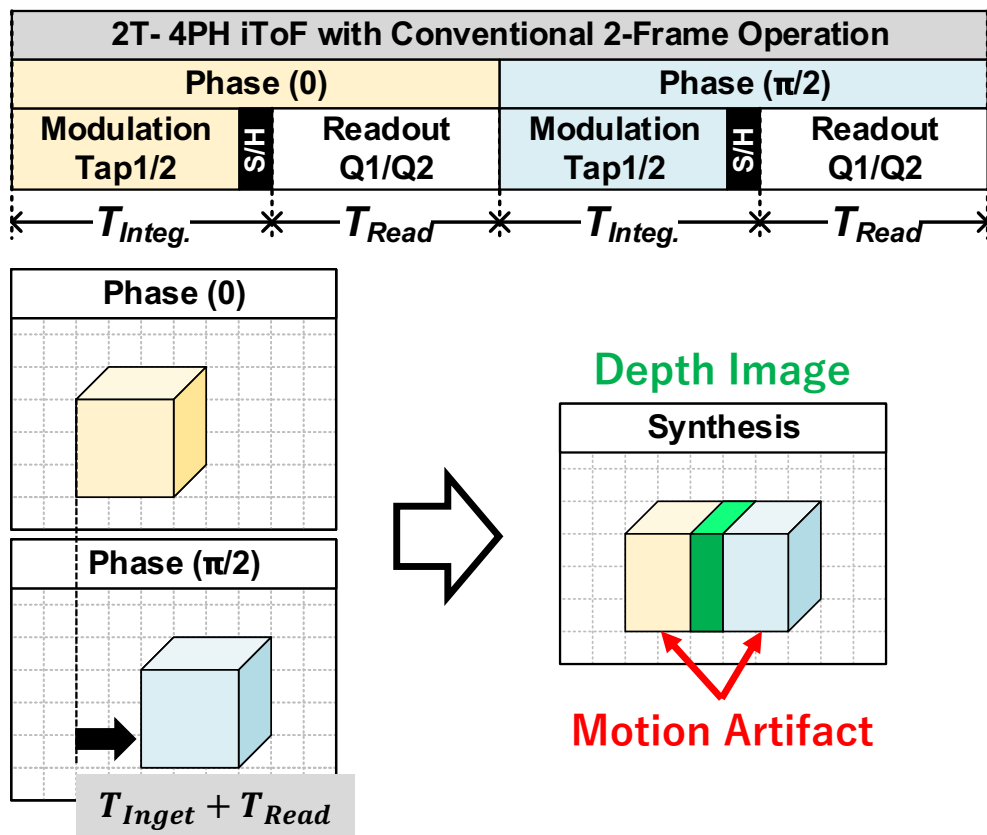
Fig. 4-2 Motion artifact issue due to frame-to-frame processing latency.

4.2. Key Technologies

To address these issues, a new approach using half-pulse (HP) modulation with sub-frame 2-tap 4-phase operation is introduced.

4.2.1. Half Pulse 2-Tap Modulation

Fig. 4-3 shows the operation principle of proposed 2T-4PH operation with half-pulse (HP) modulation. The modulation cycle time and pulse width are denoted by T_C and T_{HP}, respectively, where $T_{HP} = T_C/4$. The distance (d) can be obtained by Eq. 4-1 and Eq. 4-5.

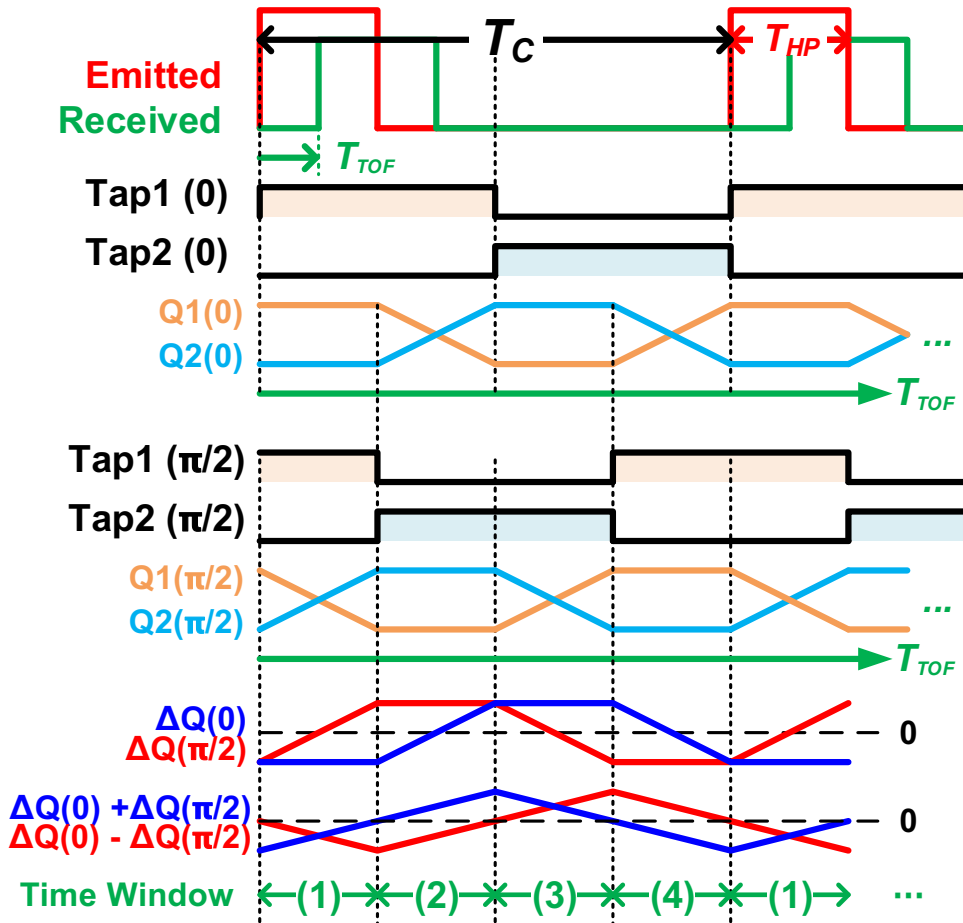$$T_{ToF} = T_{HP} \times R_{HP}, where \ 0 \le R_{HP} < 4 \qquad \text{Eq. 4-5}$$

Fig. 4-3 Operation principle of proposed HP modulation.

Here, R_{HP} is the time shift ratio defined from a range of 0 to 4 through the time-window, TW(1) to TW(4). Based on the ranging result, Eq. 4-6 and Eq. 4-7 are applied to determine the TW.

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

Eq. 4-6

$$\begin{aligned} SQ &= \Delta Q_0 + \Delta Q_{\pi/2} \\ DQ &= \Delta Q_0 - \Delta Q_{\pi/2} \end{aligned}$$

Eq. 4-7

HP modulation with two ranging methodologies, denoted as HP1 and HP2 and expressed in Eq. 4-8 and Eq. 4-9, respectively, are proposed to obtain the time shift ratios, R_{HP1} and R_{HP2}, respectively.

$$R_{HP1} = \begin{cases} 0 + \dfrac{Q2_{\pi/2} - Q2_0}{Q1_{\pi/2} + Q2_{\pi/2} - 2 \cdot Q2_0} & @\ TW(1) \\[4mm] 1 + \dfrac{Q2_0 - Q1_{\pi/2}}{Q1_0 + Q2_0 - 2 \cdot Q1_{\pi/2}} & @\ TW(2) \\[4mm] 2 + \dfrac{Q1_{\pi/2} - Q1_0}{Q1_{\pi/2} + Q2_{\pi/2} - 2 \cdot Q1_0} & @\ TW(3) \\[4mm] 3 + \dfrac{Q1_0 - Q2_{\pi/2}}{Q1_0 + Q2_0 - 2 \cdot Q2_{\pi/2}} & @\ TW(4) \end{cases}$$

Eq. 4-8

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

Eq. 4-9

Using HP modulation, HP1 method requires three individual integrated signals to obtain R_{HP1} in each TW, whereas HP2 method uses two differential demodulated charge signals for the BGLC depth calculation.

To compare the depth characteristics of the ranging methods using both SP and HP modulations, Fig. 4-4 illustrates the parameters of an iToF system. The theoretical depth noise (σ_d) can be derived by applying the propagation of errors as follows,

$$Depth\ Noise\ (\sigma_d) = \frac{c}{8f_m} \times \frac{\sigma_R}{DC} \qquad\qquad \text{Eq. 4-10}$$

$$\sigma_{R\_SP} = \frac{\sqrt{(N_S + 2N_B + 2RN^2)(1 - 4R_S + 8R_S^2)}}{N_S} \qquad\qquad \text{Eq. 4-11}$$

$$\sigma_{R\_HP1} = \frac{\sqrt{N_{ToF}(1 - R_S) + 2(N_B + RN^2)(1 - 3R_S + 3R_S^2)}}{N_S} \qquad \text{Eq. 4-12}$$

$$\sigma_{R\_HP2} = \frac{\sqrt{(N_S/2 + N_B + RN^2)(1 - 2R_S + 2R_S^2)}}{N_S} \qquad\qquad \text{Eq. 4-13}$$
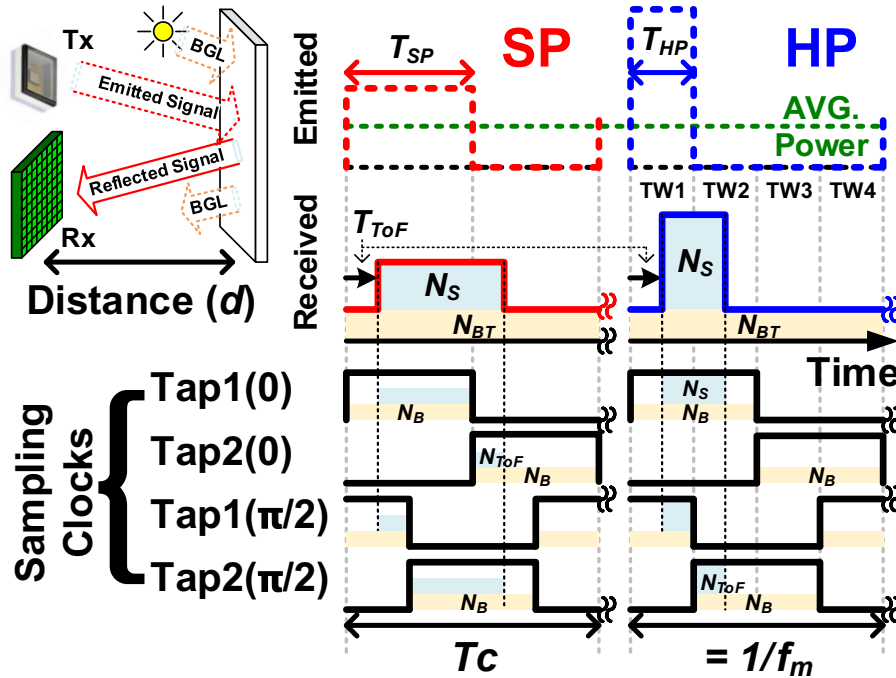

Fig. 4-4 Indirect ToF system with 2T-4PH operation using SP and HP modulation.

where f_m is modulation frequency, 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}, respectively. N_{ToF} is the number of demodulated electrons from N_S in the single tap. The readout noise referred to pixel floating diffusion (FD) is denoted by RN.

The derivation assumes that the waveforms are perfect square shapes and that the BGL has a constant value during the modulation. Therefore, R_S will increase proportionally from 0 to 1 along the time shift, T_{ToF}, in each time window, and N_B in both taps will be equal to $N_{BT}/2$. Note that the amplitude of the emitted light in HP modulation is doubled compared to SP modulation to maintain an equal average light intensity during modulation.
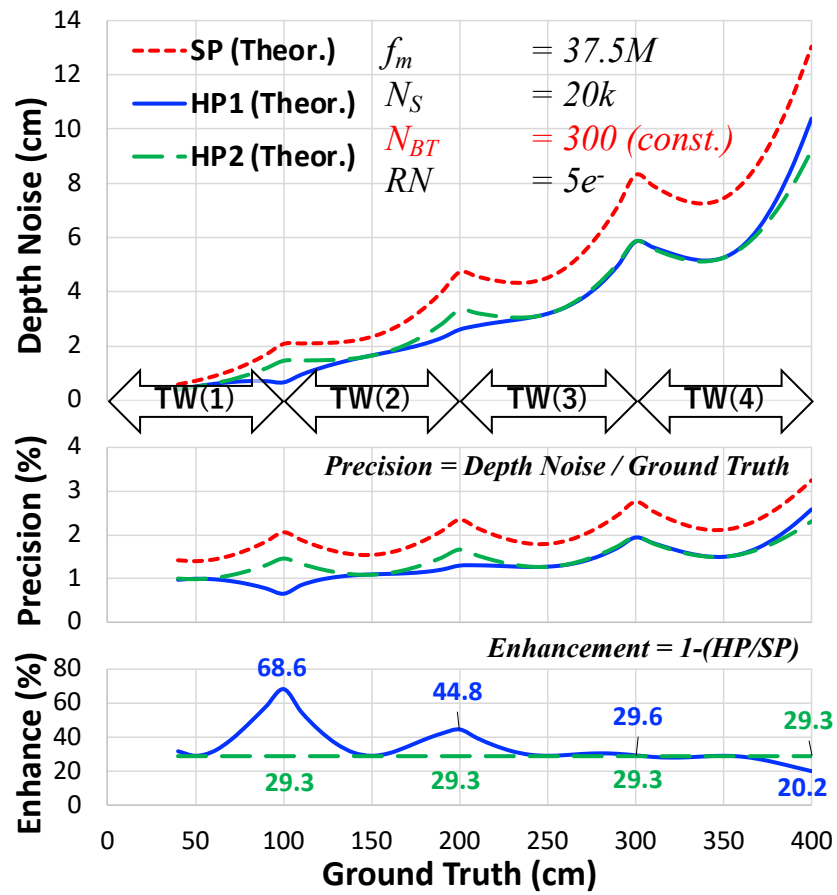
Fig. 4-5 (a) shows the SNR characteristic of an iToF system with N_{BT} of 300, RN of $5e^-$, and N_S of 20,000 at 0.4m, calculated using Eq. 1-3. Note that N_S decreases rapidly as a function of distance squared, whereas N_{BT} remains constant for all ranges.

Fig. 4-5 (b) compares the theoretical depth noise curves of SP, HP1 and HP2 ranging methods with a DC of 90%. The detection range is set to 0.4-4m at a f_m of 37.5-MHz. The precision enhancement was obtained by the ratio of HP to SP modulation and is depicted in terms of percentage.

As illustrated by the simulation results, both HP methods achieve a lower depth noise, but with a different area of strength. While the system SNR is dominated by the shot noise from N_S, HP1 method can enhance the depth precision efficiently, especially when R_S is close to 0 or 1 at each TW. On the contrary, HP2 method shows its advantage at longer distances, where $N_S < N_{BT} + 2RN^2$, and guarantees a consistent noise reduction capability for all ranges.
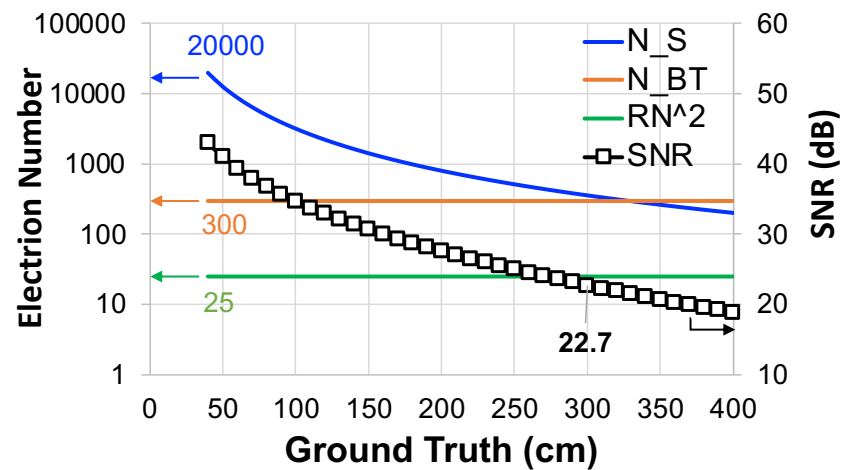
(a)

(b)

Fig. 4-5 (a) SNR characteristic of an iToF system
(b) Theoretical depth noise curves of SP, HP1 and HP2 ranging methods.

4.2.2. Sub-Frame 2T-4PH Operation

The BGLC scheme in a 2-tap iToF structure requires two exposure periods. In Section 2.3, the sub-frame ToF operation with in-pixel memory was introduced, which can minimize the deadtime between subframes due to the short sample/hold (S/H) period.

Fig. 4-6 illustrates the concept of motion artifact suppression with sub-frame 4-phase sampling. Compared to the conventional 2-frame operation depicted in Fig. 4-2, the sub-frame operation provides a reduction in frame-to-frame latency owing to the compact modulation periods. This allows an effective suppression of motion artifacts while capturing moving targets. In the sub-frame 2T-4PH operation, the array readouts are performed after finishing the 4-phase modulation.
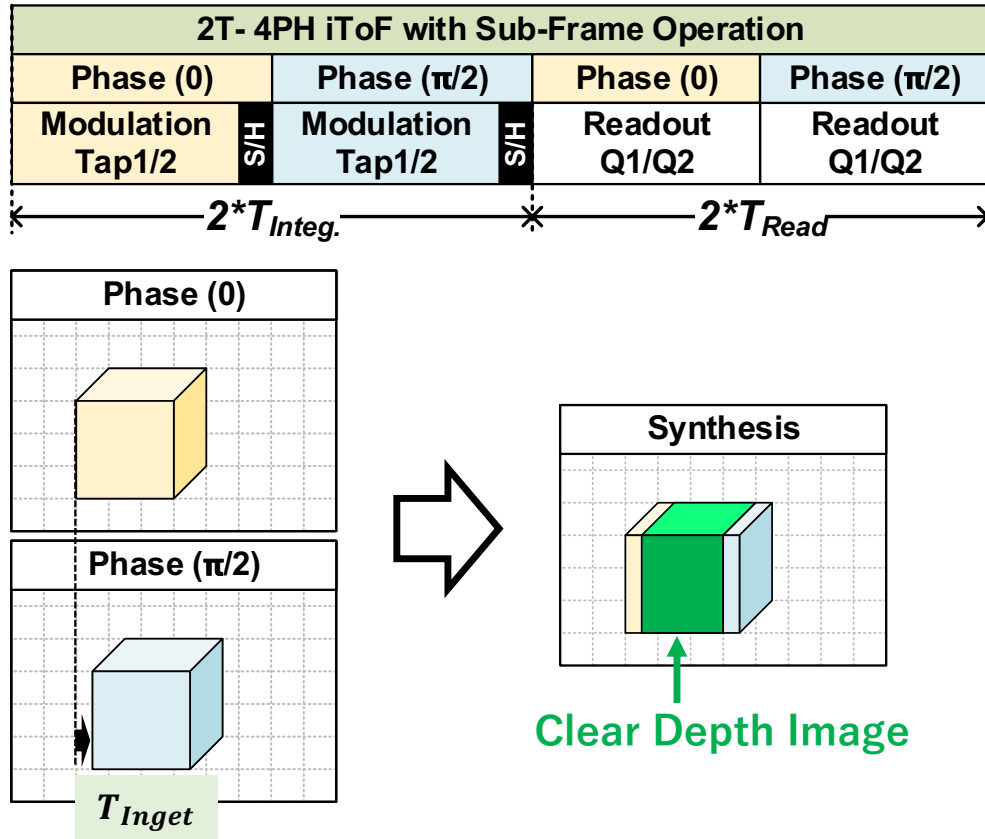
Fig. 4-6 Sub-frame 2T-4PH operation for motion artifact suppression.

4.3. Equivalent System

The proposed technologies were verified using the 4-tap iToF imager in Section 3.1 with adjusted operation timing.

4.3.1. Pixel Structure

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 2×8 1-T 1-C analog memory (C_{MEM}) with control devices (MW, MRST, MTs, MSs), which share a column readout buffer (MSF).
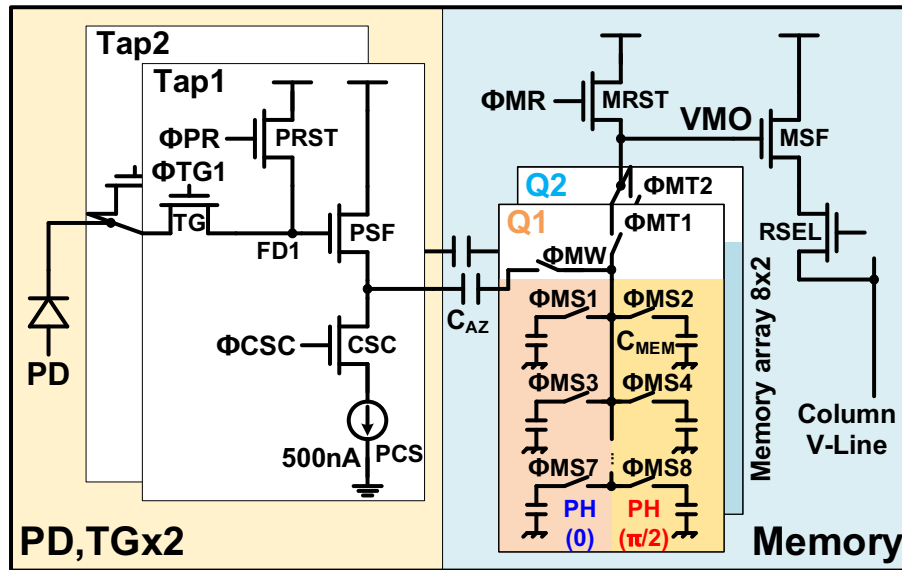
Fig. 4-7 The equivalent 2-tap pixel structure with in-pixel memory array.

4.3.2. Sensor Operation

Fig. 4-8 (a) depicts the timing diagram of the sub-frame 2T-4PH integration, where $PH(0, \pi/2)$ modulations are performed and sampled into the memory array alternatively. Auto-zeroing sampling is carried out in each subframe to eliminate the reset thermal noise.

Fig. 4-8 (b) depicts the timing diagram of array readout. The charge-domain binning is applied to each phase by mixing the signal charges in the corresponding memories before the column sampling to increase the system SNR. Lastly, the sampled signals of $PH(0)$ and $PH(\pi/2)$ are readout in sequence.
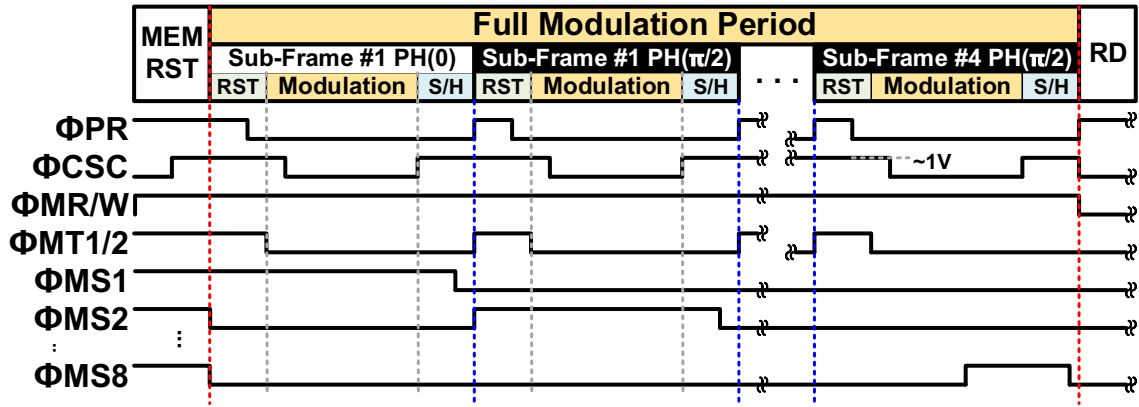
(a)
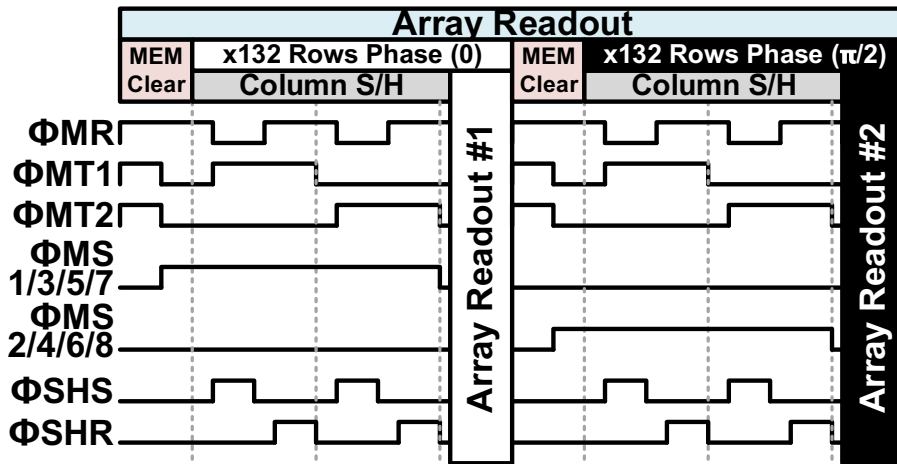
(b)

Fig. 4-8 Timing diagram of (a) Sub-frame 2T-4PH operation (b) Array readout.

4.4. Verification

The basic sensor characteristics has been summarized in Section 3.3, where a single-tap FWC around 12ke^-, and a readout noise floor of 5.4e^- with 4-subframe averaging were confirmed. A DC of 85% with a 20 ns demodulation pulse width was obtained, where the modulation light was generated by an 850 nm VCSEL.

4.4.1. Depth Accuracy and Precision

For the range measurement, each subframe has an exposure time of 0.6ms, and the imaging system was set to 60 fps with the modulation frequency of 25 MHz. This corresponds to an unambiguous range of 1.5 m for a single time-window (TW). However, due to the circuit timing constraints, the measurable range was limited to 1.5 m with conventional SP modulation using TW(1), and 3.0 m with proposed HP modulation using TW(1) and TW(2). The measured depth nonlinearity was <2%.

The experimental data in Fig. 9 shows that the depth precision was enhanced by 44% and 31% at 1.5 m using the HP1 and HP2 ranging method, respectively. The experimental results demonstrate a good agreement with the theoretical calculations presented in Section 4.2.1, which can be utilized to estimate the depth precision for the 2-tap iToF ranging methods. The differences observed between the experimental and theoretical results were explained in Section 3.3.4. The simulation results show that HP1 and HP2 can reduce noise by >25% and >29%, respectively, for the range of 0.4-3 m.
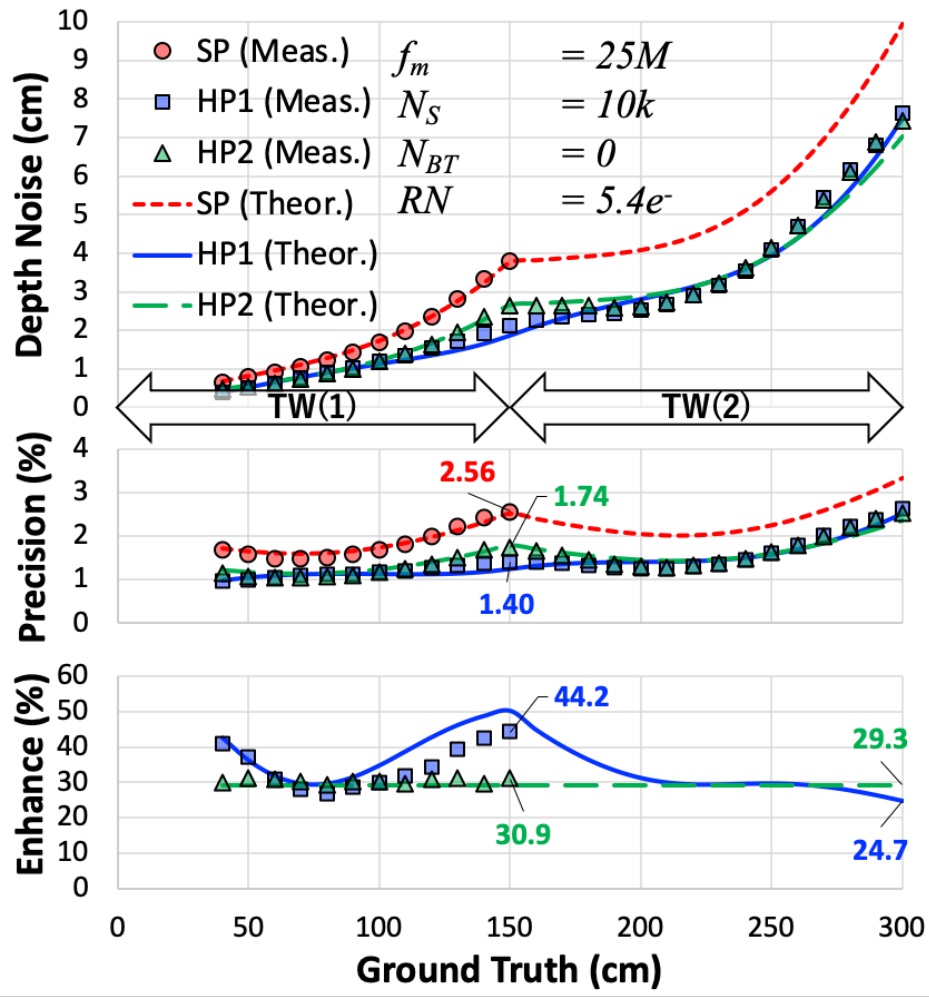
(a)
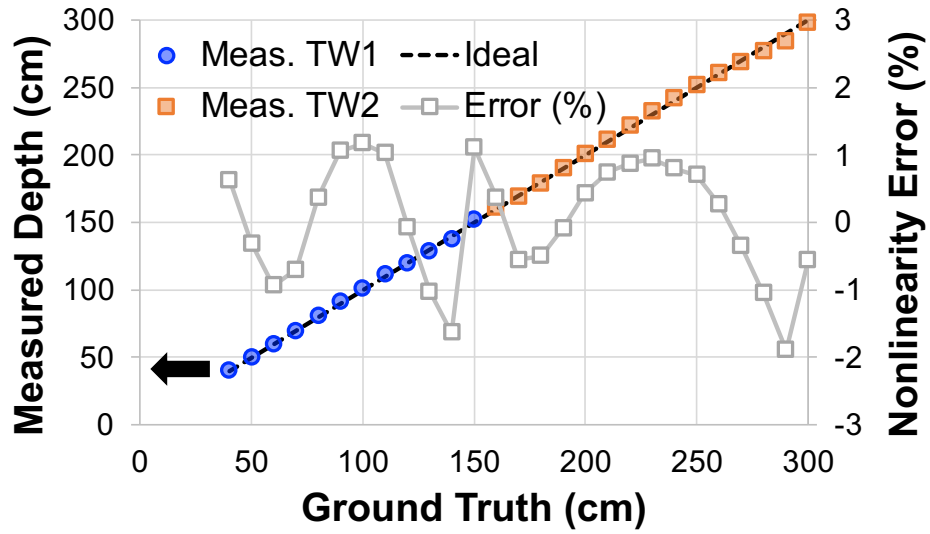
(b)

Fig. 4-9 (a) Depth precision (b) Depth accuracy.

4.4.2. Sample Images

Fig. 4-10 demonstrated the captured sample images with a hand waving from the left-front to the right-rear. A serious depth distortion was observed while using the conventional 2-frame operation with SP modulation. In contrast, with the help of proposed HP ranging methods and sub-frame sampling, clearer depth images were obtained with lower noise and suppressed motion artifact.

It should be noted that there is still some depth uncertainty at the edge of the hand in the image captured using the HP2 method, which uses both ΔQ_0 and $\Delta Q_{\pi/2}$ for the depth calculation. This uncertainty is caused by the multiple subframes with long integration time that increased the time differences of $PH(0, \pi/2)$. To further improve the motion artifact, the emitter power can be increased or IR-sensitivity enhancing technologies such as micro lens, deep trench isolation (DTI), and pyramid surfaces for diffraction (PSD) structure can be adopted to reduce the required exposure time.
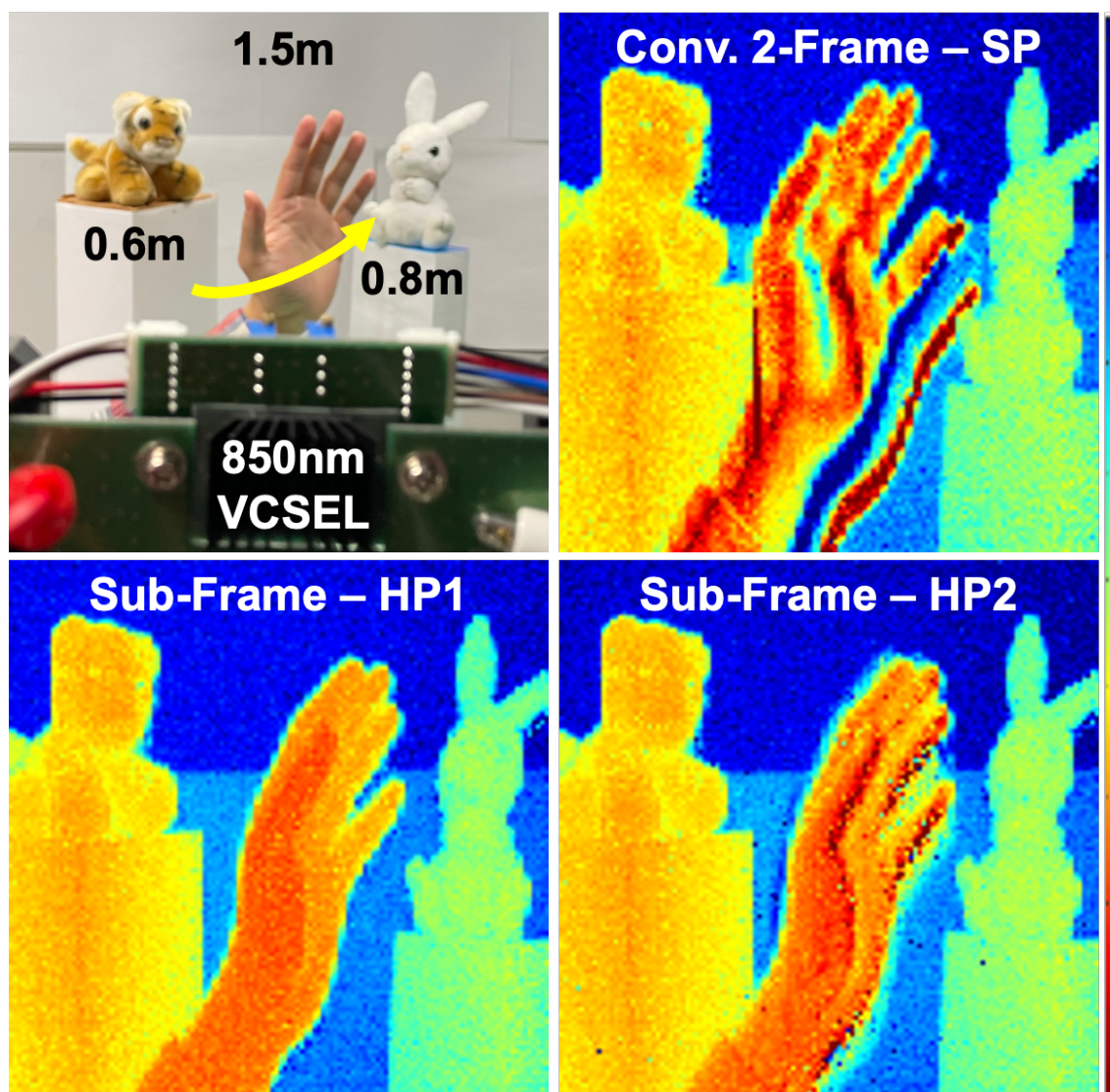
Fig. 4-10 Captured sample images with hand movement.

4.5. Discussions

The HP ranging methods enable enhancement of the depth precision for a 2-tap range imager, without requiring an increase in modulation period or frequency. With the use of HP modulation, the HP1 and HP2 ranging methods offer advantages under different circumstances.

Fig. 4-11 (a) shows the simulated depth noise curves at a f_m of 37.5 MHz without BGL. Here, a fixed N_S/RN of 5000 was used for different capacitance values of FD, which is inversely proportional to the conversion gain. While targeting a depth precision around 1% error for 0.4-4 m range, HP1 method can achieve with the lowest amount of signal charge ($N_S = 20$k), whereas the SP modulation needs 4 times higher number to reach a similar performance. This indicates a higher framerate and lower power consumption can be expected using the HP modulation.

Therefore, the HP1 method is recommended for indoor 3-D imaging applications, such as industrial automation or motion sensing devices, where high precision and high framerate are desired for a 2-tap iToF range imager with high system SNR and weak ambient light influence.

On the other hand, in outdoor environments where there is stronger ambient light, HP2 method can offer a more stable depth imaging performance. As shown in Fig. 4-11 (b), HP2 provides a better depth precision at longer distances compared to SP method, which requires double the amount of signal charge. Meanwhile, the implementation of the on-chip BGLC scheme can be achieved through different methods, such as cross-connected capacitors or a charge amplifier with capacitive feedback. The differential output can also help to reduce data-rate burden and power consumption for high-speed readout circuits, leading to higher resolution or higher framerate. Additionally, the

common-mode rejected Δ-INT BGLC scheme can enable sub-integration accumulation of the modulated signal while preventing the FD from saturation under strong BGL. It has been reported that this scheme can tolerate ambient light levels over 100 klx.

The HP2 ranging method can be introduced to a wider range of applications, such as outdoor robotics and autonomous vehicles, owing to the outdoor usage capability.
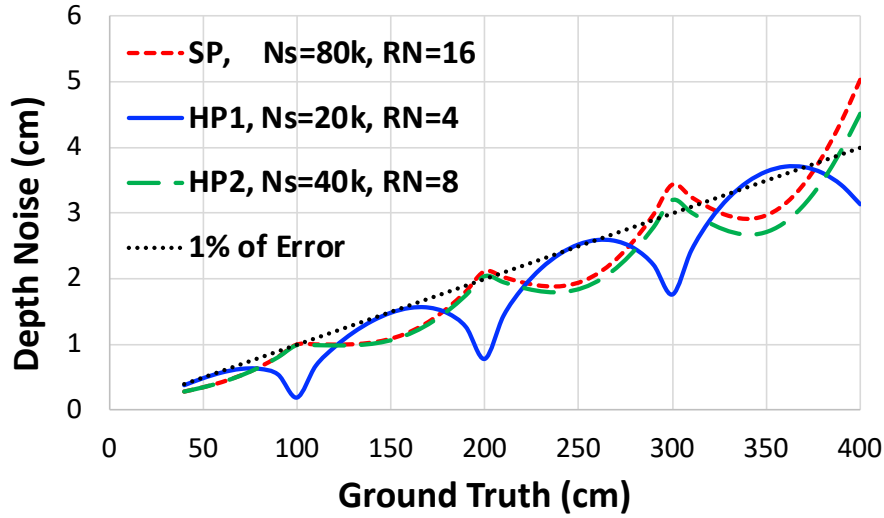
(a)
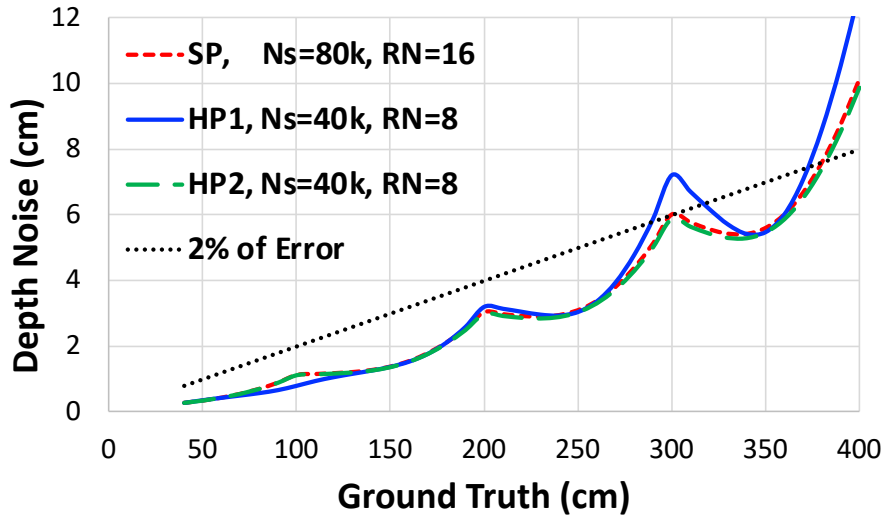
(b)

Fig. 4-11 Estimated sensor performance using SP and HP modulation
(a) without BGL ($N_{BT} = 0$) and (b) with a constant BGL ($N_{BT} = 2000$)

4.6. Summary of Chapter. 4

In this chapter, a 2-tap 4-phase (2T-4PH) iToF ranging method using half-pulse (HP) modulation with sub-frame operation were presented.

The HP modulation is achieved by applying a modulated light with half of the pulse width and double the amplitude. By applying different ranging algorithms, HP1 and HP2 BGLC ranging results can be obtained. Both HP1 and HP2 methods have been shown to provide better ranging precision than continuous square-pulse (SP) modulation and be utilized in different ways. The HP1 method is recommended for indoor applications where the highest depth precision is desired. In contrast, the HP2 method has an advantage in lower SNR ranging systems, such as those in strong ambient light environments or for long-distance ranging.

On the other hand, the combination of the 2T-4PH BGLC scheme and sub-frame operation effectively suppresses motion artifacts due to the compact modulation periods.

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

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

Chapter. 5.

Proposal of a BSI 3D-Stacked 2-Tap iToF Range Imager

5.1. Key Concepts

The prototype iToF sensor demonstrated in Chapter. 3 exhibits the potential to achieve high quality depth images while also having the capacity to provide high temporal resolution 3-D images. However, to enable more applications, longer recording durations of burst images are desired, which requires a larger number of in-pixel memory cells. This, in turn, leads to larger pixel sizes and lower fill-factors, resulting in lower resolution and framerate. On the other hand, although the 4-tap iToF ranging scheme has the advantage of mitigating motion artifact and achieving lower depth noise, the requirement of more transistors, the potential for tap mismatch, and the complexity of the pixel layout may limit the performance of sensor system.

To address these issues, the introduction of a 2-tap iToF pixel structure, backside illumination (BSI), and 3D stacking technologies can significantly improve pixel area efficiency while realizing the next generation of high-speed and high-precision range imager. Additionally, the depth noise and motion artifact can be suppressed using the proposed half-pulse (HP) 2-tap 4-phase (2T-4PH) ranging method with sub-frame operation, which was introduced in Chapter. 4.

5.2. Development of Charge Modulator System

Fig. 5-1 shows the circuit diagram of the proposed 2-tap 4-phase iToF pixel with a size of 16.8^H μm×16.8^V μm. The pixel is constructed by a 3-D stacked structure with sensor chip and memory chip. The 2-tap charge modulator is implemented in sensor chip, which comprises a BSI high-speed charge collection photodiode (PD), demodulation gates (TGs), two sets of source follower (PSF) and current source (PCS) with cascode switch (CSC). The memory chip consists of two auto-zeroing capacitors (C_{AZ}), 4×16 1-T 1-C analog memory array, control devices (MW, MRST, MTs), and voltage buffers (MSF, RSEL) for column readout. Note that two column buffers (MSF) are implemented to increase the column readout speed. The 3-D stacking nodes between chips connects the output of PSF and the bottom plate of C_{AZ}.
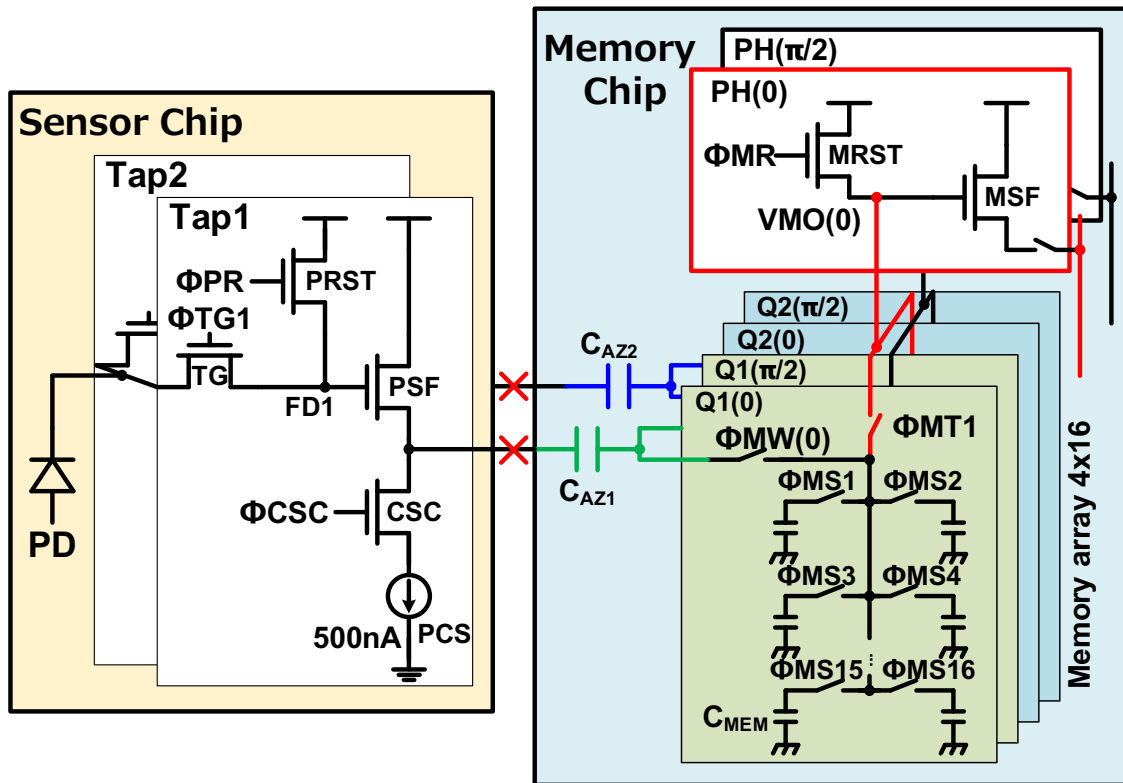
Fig. 5-1 The circuit diagram of proposed 2T-4PH iToF pixel.

➢ **Sensor Chip**

Fig. 5-2 (a) shows the layout diagram of the 16.8-μm pitch pixel, while Fig. 5-2 (b) depicts the cross-sectional diagram of the BSI photodiode (PD).

The design concepts are summarized as follows.

1) To generate the lateral and vertical E-fields that collect the photon-electrons to the center of the modulator efficiently, three levels of n-layer (N1-N3) and a deep n-well (DNW) are used. This approach is applied to a 6-μm-thick P-epitaxial on N-substrate wafer.

2) To reduce photon-charge crosstalk and stabilize the substrate ground voltage, the n-type and p-type pickups are placed surrounding the PD.

3) To enhance the photo-responsivity for IR wavelengths, which have a deeper absorption length, the bottom metal (Metal 1) covers the PD area to reflect the light comes from the backside. The metal is connected to ground voltage.

(a)   (b)
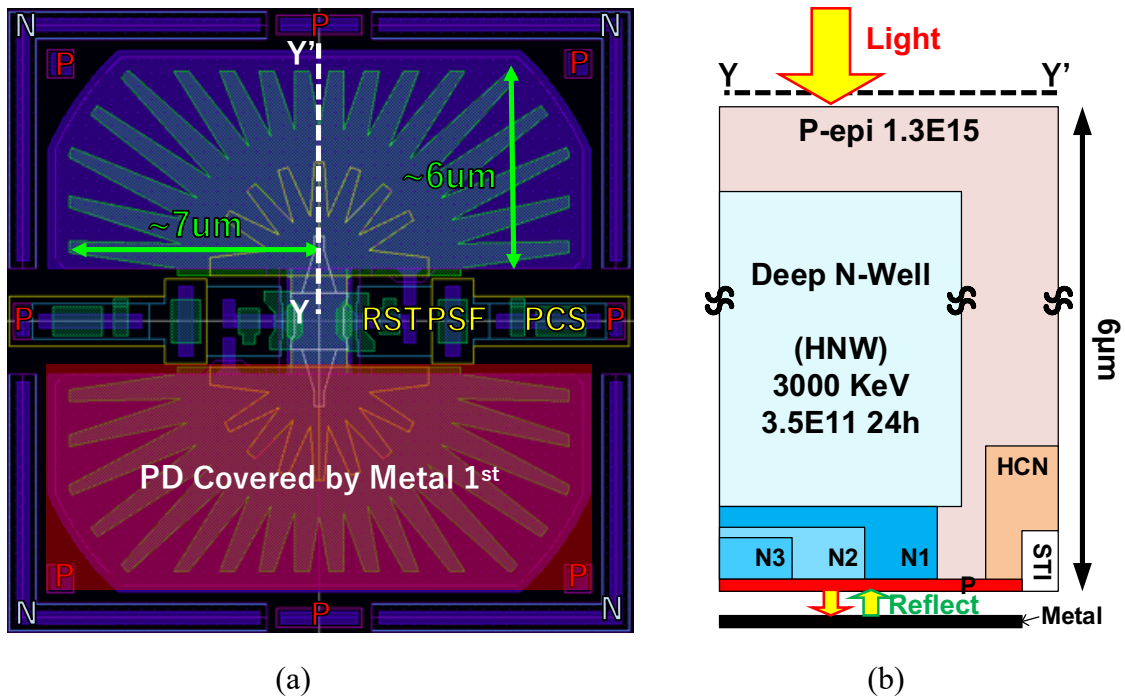
Fig. 5-2 (a) Layout diagram (b) Cross-sectional diagram of the BSI 2-tap iToF pixel

Fig. 5-3 (b), (c) and (d) show the simulation results of the charge demodulation with potential diagram at X-Y, Y-Z, and X-Z plots, respectively. In this simulation, an electron was initially placed at a depth of 5 μm from the far end of the PD.

The results show that the electron was successfully transferred to the FD along the created potential gradient, where the vertical E-field guided the electron towards the surface of the PD and the lateral E-field collected the electron to the demodulating gate. The designed charge modulator can collect the electrons within 1.0 ns from the backside.
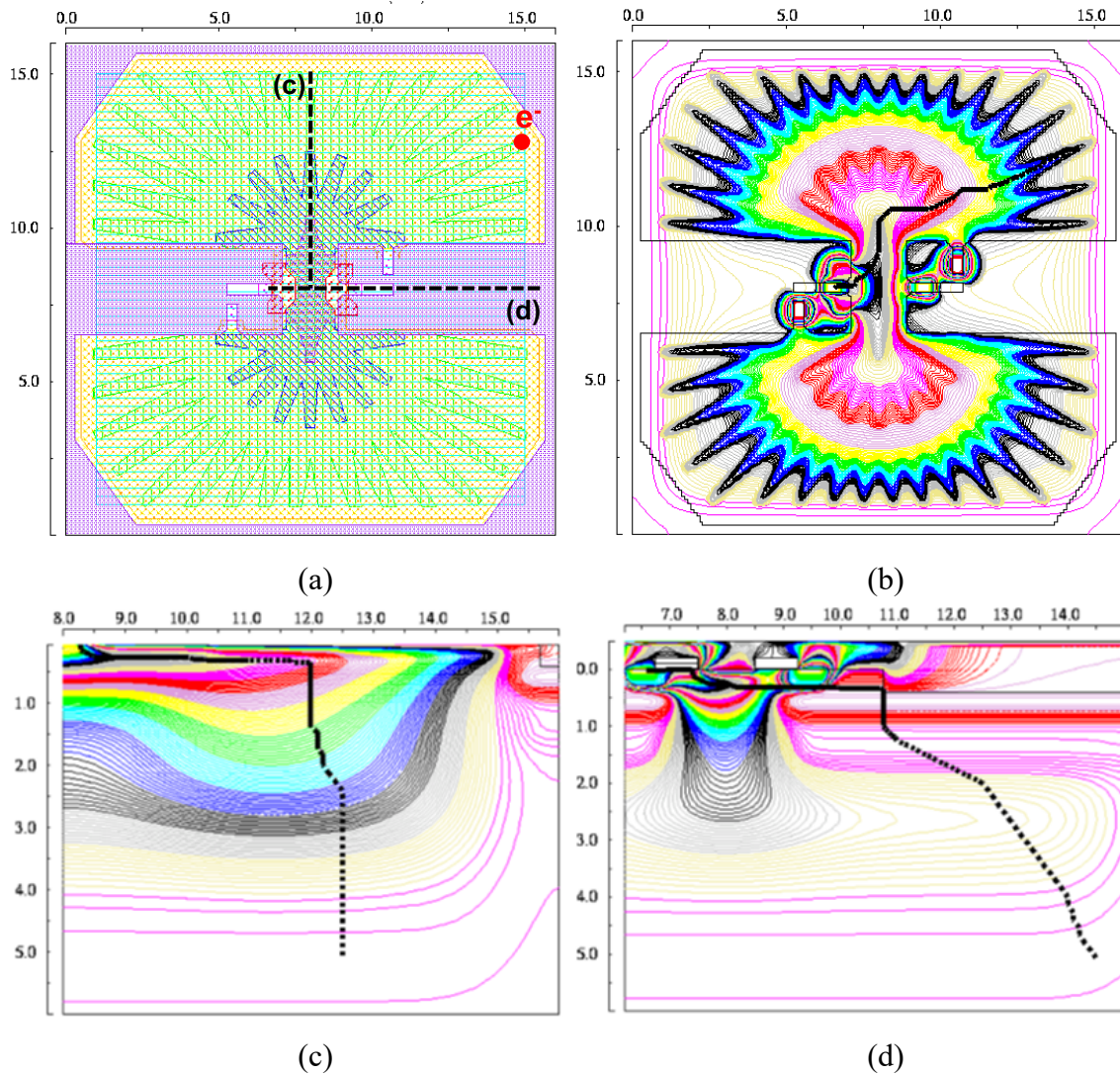
Fig. 5-3 Simulated potential diagrams with electron transfer path.
(a) Layer plot (b) X-Y plot (c) Y-Z plot (d) X-Z plot

➢ **Memory Chip**

Fig. 5-4 shows the simplified circuit and layout diagram of the 16.8-µm pitch 4-phase memory array, which is composed of 16 memory cells for each phase, and two C_{AZ} for 2-tap modulation with auto-zeroing operation. In this design, the estimated capacitance of C_{MEM} and C_{AZ} are around 25 fF and 200 fF, respectively.

Note that the memory array of $Q1(0)$ and $Q2(0)$ share the same MSF buffer to eliminate the offset induced by process variation. Therefore, using either Eq. 4-3 or Eq. 4-9 for the BGLC ranging calculation can result in better depth accuracy. The 3D stacking nodes, serving as the bottom plate of the C_{AZ}, are connected to top metal (Metal 5) denoted by the red "X".
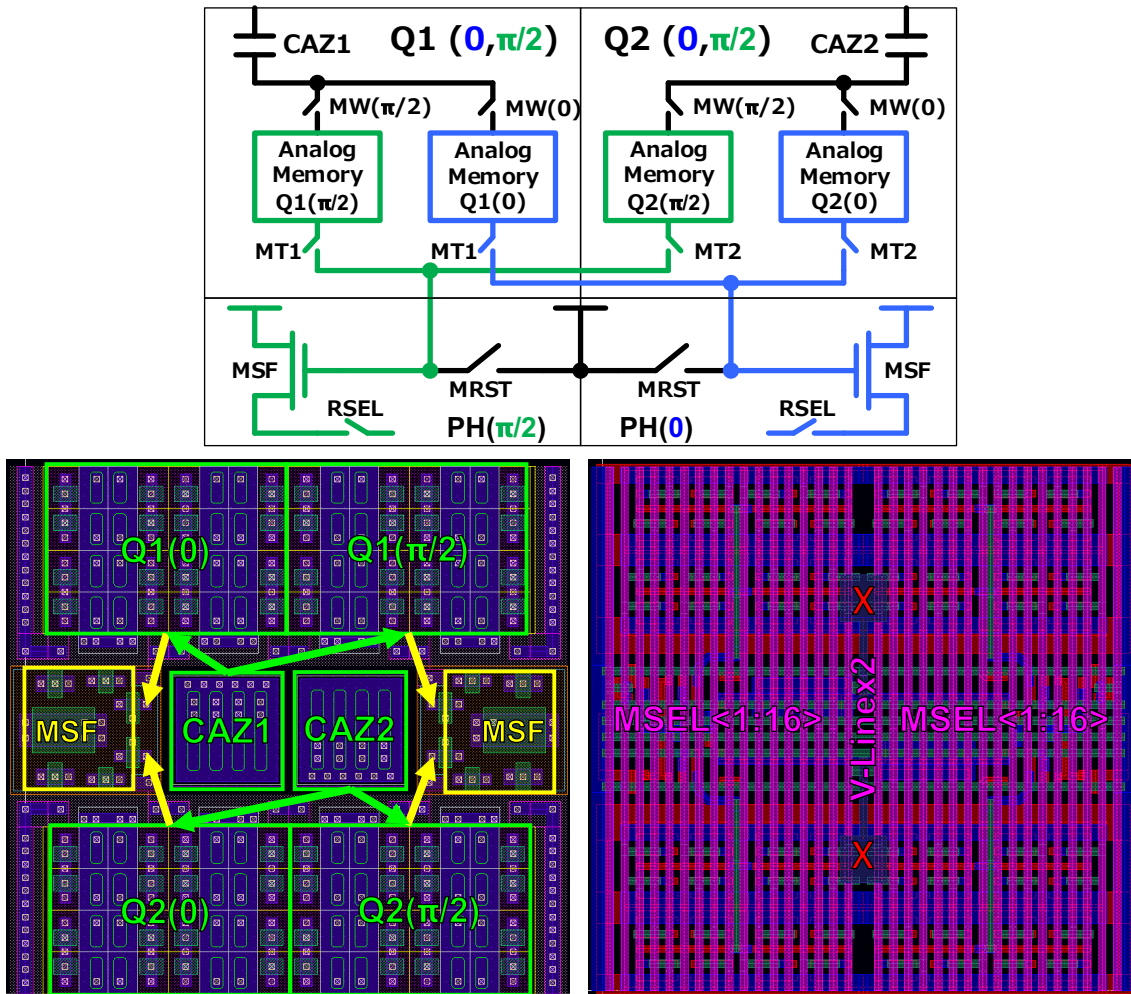
Fig. 5-4 Simplified circuit and layout diagram of the 4-phase memory array.

➢ **2-Tap TG Clock Generator**

To ensure good linearity and ranging accuracy, it is crucial to carefully control the duty ratio between TG1 and TG2, while also avoiding overlap between the TG clocks, which can lead to a reduction in demodulation contrast (DC).

Fig. 5-5 illustrates the system diagram of TG clock generator and the generated waveforms. The clock divider utilizes D flip-flops (DFFs) to ensure a duty ratio of 50%. The non-overlap circuit produces TG clocks with a latency of ~0.5 ns between clocks to avoid overlap. The switching of $PH(0, \pi/2)$ can be controlled by the TG_PH signal. Finally, the TG_Flip signal can be used to flip the TG clocks as a potential solution for balancing tap mismatch.
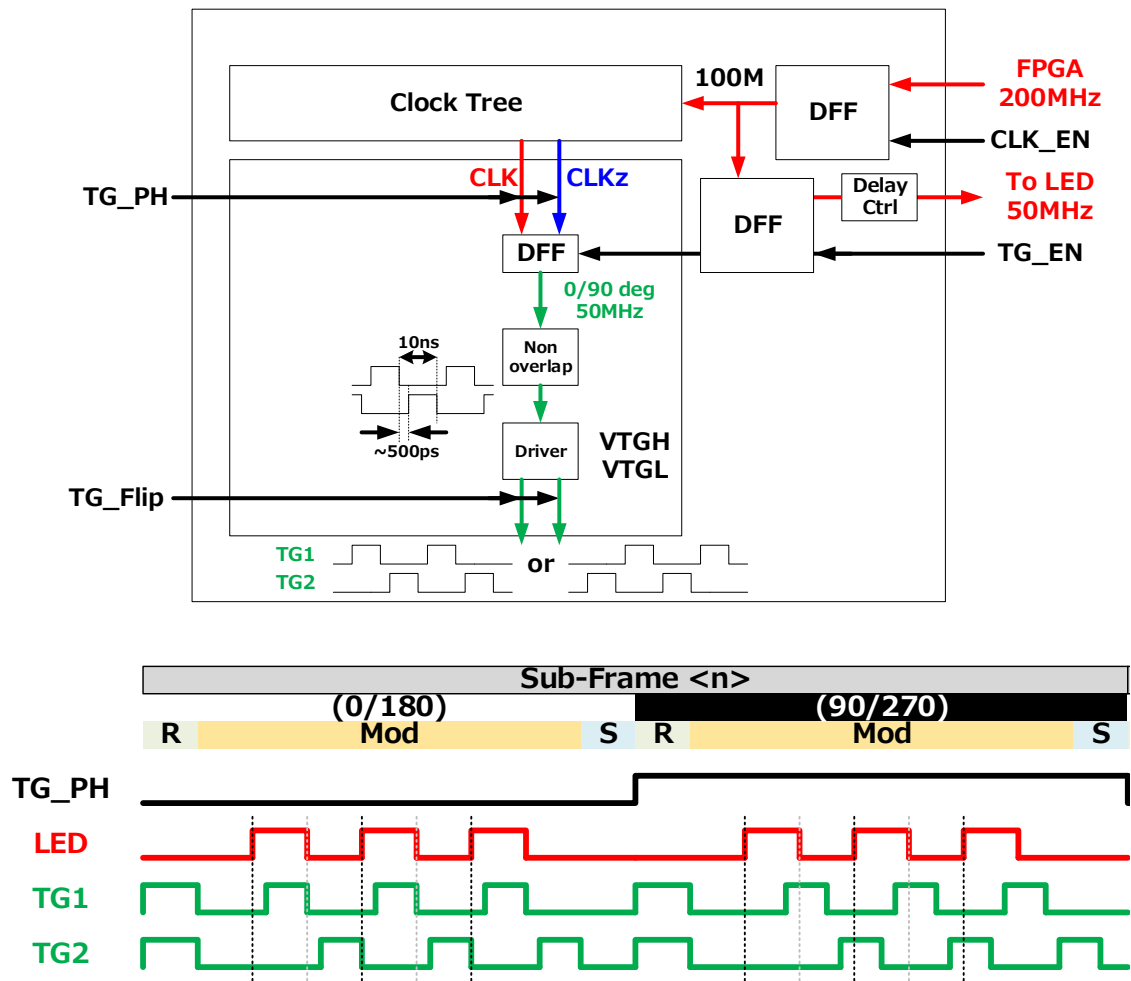
Fig. 5-5 System diagram and waveform of 2-tap TG clock generator.

The circuit simulated waveforms is shown in Fig. 5-6 (a), where a 200-MHz CLK_IN clock from FPGA was used. The CLK_ENH enables the DFF to generate the divided clock of CLK_DIV2 at 100-MHz. The TG_ENH signal is used to enable the TG phase generator, while the PIX_RST signal is used to force the TGs to reset the pixel FDs.

Fig. 5-6 (b) and (c) show the zoom-in views of the PH(0) and PH($\pi/2$), respectively, which are controlled by TG_PH. The generated pulse width of ~10ns and the nonoverlap TG clocks of ~550ps were confirmed.
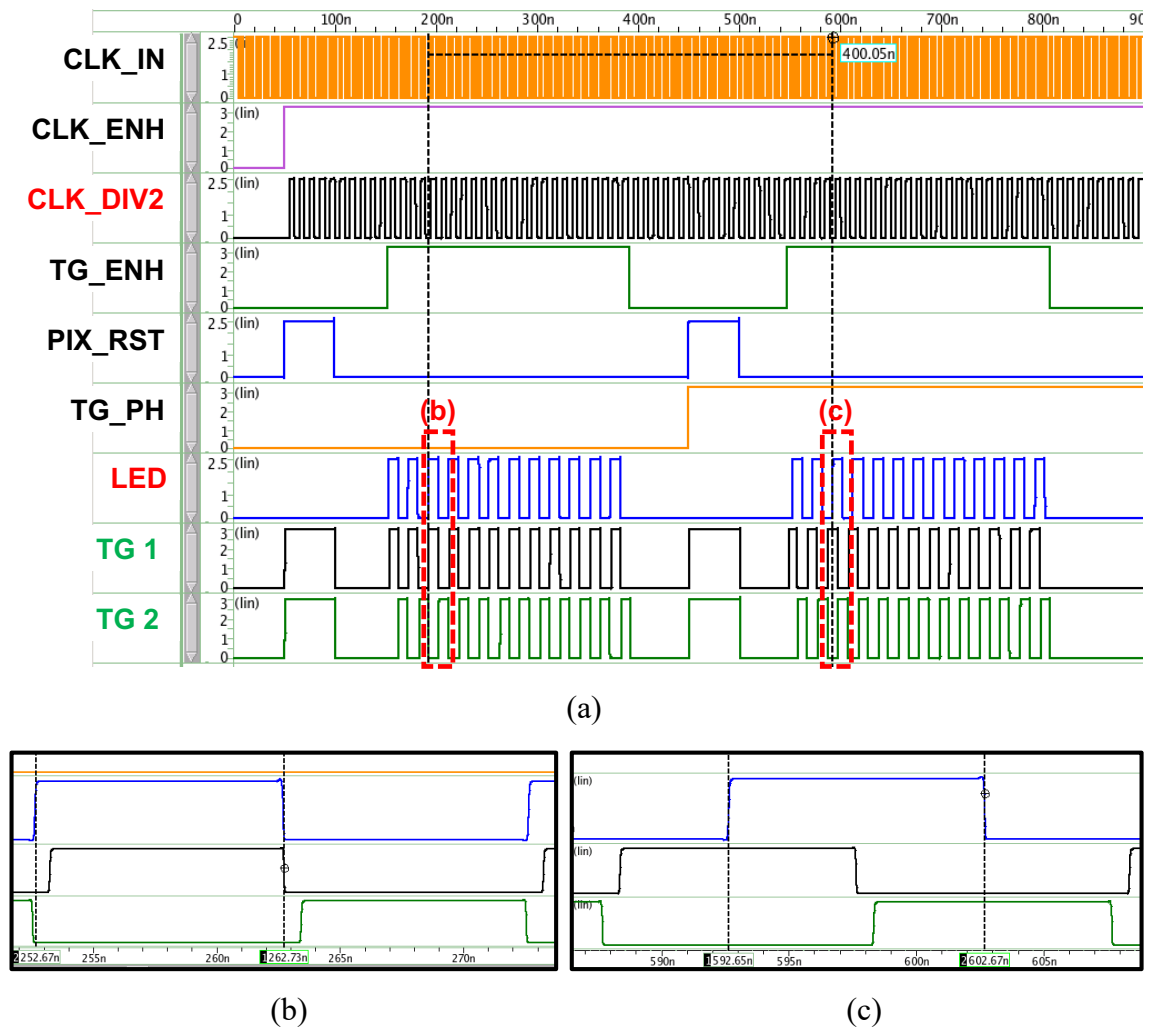
Fig. 5-6 Circuit simulation of the TG clock generator.
(a) Signal waveforms (b) Phase 0 (c) Phase $\pi/2$

5.3. Comparison

The dimensions of the prototype 4-tap (Section 3.1.2) and the proposed 2T-4PH iToF pixels are compared in Table 5-1.

Firstly, the backside illumination (BSI) technique enables an ideal fill-factor (FF) of 100%, which enhances the efficiency of charge collection. The deep n-well structure increases the depth of the photodiode with a strong vertical E-field, reducing the charge transfer time and improving the demodulation contrast (DC). Additionally, although a thinner epitaxial thickness is used, metal reflection doubles the equivalent absorption length, enhancing the photon-responsivity for infrared light. These benefits improve the depth noise, reduce the required exposure time and power consumption.

Secondly, the 3D stacking technology significantly improves area efficiency. Furthermore, the 2-tap pixel requires only half the number of auto-zeroing capacitors (CAZ), enabling doubling of the memory cell count in the pixel implementation. The larger memory array allows for better noise performance under HP mode and longer record length under HS mode. Additionally, the use of two column voltage buffers allows for parallel readout, which helps increase the frame speed.

Finally, he designed 4-phase TG clock generator with non-overlap clocks and flip function for 2-tap operation can enhance the depth accuracy with improved linearity.

Using the proposed 2T-4PH iToF pixel with half-pulse modulation and sub-frame operation, a higher resolution, higher depth precision, higher framerate, and longer recording length can be expected in the development of the next generation high precision and high-speed iToF range imager.

Table 5-1 Comparison of 4-tap and 2T-4PH iToF pixel.

	Parameter	Prototype 4-tap	Proposed 2T-4PH
Pixel	Process Tech.	0.18μm FSI	0.18μm BSI
	Pixel pitch (μm)	22.4 × 16	16.8 × 16.8 (Stack)
	Pixel architecture	4-tap	2-tap
Modulator	P-epi thickness (μm)	8 μm	6 μm
	Vertical E-field depth (μm)	~1.5 μm [1	~4.5 μm [1
	Maximum Charge Transfer Time	~0.8 ns	~1.0 ns
	Fill factor (%)	~21.6	100
Memory	Memory cell	4 × 8	4 × 16
	Auto-zeroing cap.	4 × 170 fF	2 × 200 fF
	SF buffer	1	2
PERF.	Read noise @ HP mode	407 μV	360 μV [2
	Record Length @ HS mode	8 frames w/ BGLC 8 frames w/o BGLC	16 frames w/ BGLC 32 frames w/o BGLC

[1]Spectra simulation (>1000 V/cm); [2]Estimated by circuit stage noise (Fig.3-7);

Chapter. 6.

Conclusion

This dissertation has presented a comprehensive study of a high precision and high-speed indirect time-of-flight (iToF) CMOS image sensor, including background knowledges, key technologies, system development details, potential applications. Additionally, a practical implementation for the future improvements were also proposed.

In Chapter. 1, an overview of 3D imaging technologies and their key requirements was introduced. Out of these approaches, the iToF ranging methodology was selected for this study due to its advantages, such as lower power consumption, good depth precision with scalable range, small system footprint, and potential for achieving higher frame rate.

The operation principle and important parameters of the iToF ranging system were described. The targets of the iToF imager development and reported techniques have been explained, including achieving a better charge transfer efficiency, higher depth precision, wider detection range, stronger ambient light resistance, and suppressing motion artifacts. However, the current iToF imaging system designs require trade-offs when adopting these techniques, resulting in a limited field of applications. To broaden the potential applications of 3D imager in machine vision field, a prototype high precision and high-speed iToF range image sensor was developed.

Chapter. 2 introduced key technologies for realizing the targeted iToF imager, including 4-tap short pulse (SP) modulation, sub-frame ToF operation, high-speed charge modulator and high-density memory with auto-zeroing operation.

Conclusion

The 4-tap iToF scheme was implemented to reduce motion artifact for moving scenes, while SP modulation improved depth precision under a high SNR system. A high CG pixel with sub-frame ToF operation was proposed to improve SNR, with an optimal subframe number of 8 for better depth image quality. Charge domain binning reduced system noise in high precision (HP) mode, while subframe readout achieved high temporal resolution in high-speed (HS) mode. To maintain good DC and increase charge collection efficiency, a relatively large PD with optimized potential gradient was required. The in-pixel memory with high-density Si trench capacitor and auto-zeroing operation enabled sub-frame operation with a smaller memory array and higher PD fill-factor (FF).

Chapter. 3 described the details of the implementation and verification of the prototype iToF sensor with high-precision and high-speed depth imaging capability. The applicability of the proposed range imager was also demonstrated.

The proposed iToF imager features a pixel array of $134^H \times 132^V$ pixels, with a pixel size of 22.4^H μm $\times 16^V$ μm, and each pixel is equipped with a 4-tap high-speed charge modulator and a 4×8 memory array. The sub-frame ToF operation consists of 8 subframes with SP modulation. HP mode imaging adopts 4-tap modulation with memory averaging readout, while HS mode uses pseudo-2-tap operation and obtains 8 frames of burst images by readout the memory individually. The iToF imager achieved an unprecedented performance in HS mode, achieving up to 10 Kfps range imaging. Also, less than 1.77% depth noise and an R-FoM, which determined best efficient working range, of 16pJ/pixel were achieved at 0.4-5.4 m in HP mode.

The iToF imager has the potential to be used in automotive safety systems for behavior monitoring and accident detection. This represents a new direction for high-speed 3-D imaging applications in machine vision and beyond.

Conclusion

In Chapter. 4, a novel 2-tap 4-phase (2T-4PH) iToF ranging method using half-pulse (HP) modulation with sub-frame operation was introduced.

The HP modulation was achieved by using a modulated light with half the pulse width and double the amplitude, which allows obtaining different ranging algorithms for HP1 and HP2 BGLC ranging results. Both methods were demonstrated to provide better ranging precision than continuous square-pulse (SP) modulation and can be used in different scenarios. HP1 is recommended for indoor applications requiring high depth precision, whereas HP2 has an advantage in lower SNR ranging systems or long-distance ranging under strong ambient light. In addition, the combination of the 2T-4PH scheme and sub-frame operation effectively suppresses motion artifacts due to the compact modulation periods, ensuring high-quality depth imaging with moving targets.

The proposed technologies can be adopted in 2-tap iToF sensors to enhance their ranging performance without increasing the modulation period or frequency, resulting in a higher frame rate and lower power consumption.

Chapter. 5 presents the design of a practical implementation of a 2T-4PH iToF pixel, which incorporates innovative features to enhance performance. The use of backside illumination (BSI), deep n-well (DNW) structure, and metal reflection improves charge collection efficiency and demodulation contrast (DC). Additionally, the 3-D stacking technology enables a larger memory array size, resulting in longer recording duration under HS mode. Furthermore, the proposed HP modulation and sub-frame operation combination allows for higher resolution, depth precision, and framerate. These advancements hold great potential for the development of the next generation high precision and high-speed iToF range imager.

Conclusion

Finally, the value of this study is summarized in Fig. 6-1. The proposed sub-frame ToF operation is the key to enable innovations in high-speed and high precision depth imaging. Moreover, the novel 2T-4PH half-pulse ranging methods with the development of BSI with 3-D stacking modulator can lead to better sensor performance.

Fig. 6-1 Value of this study
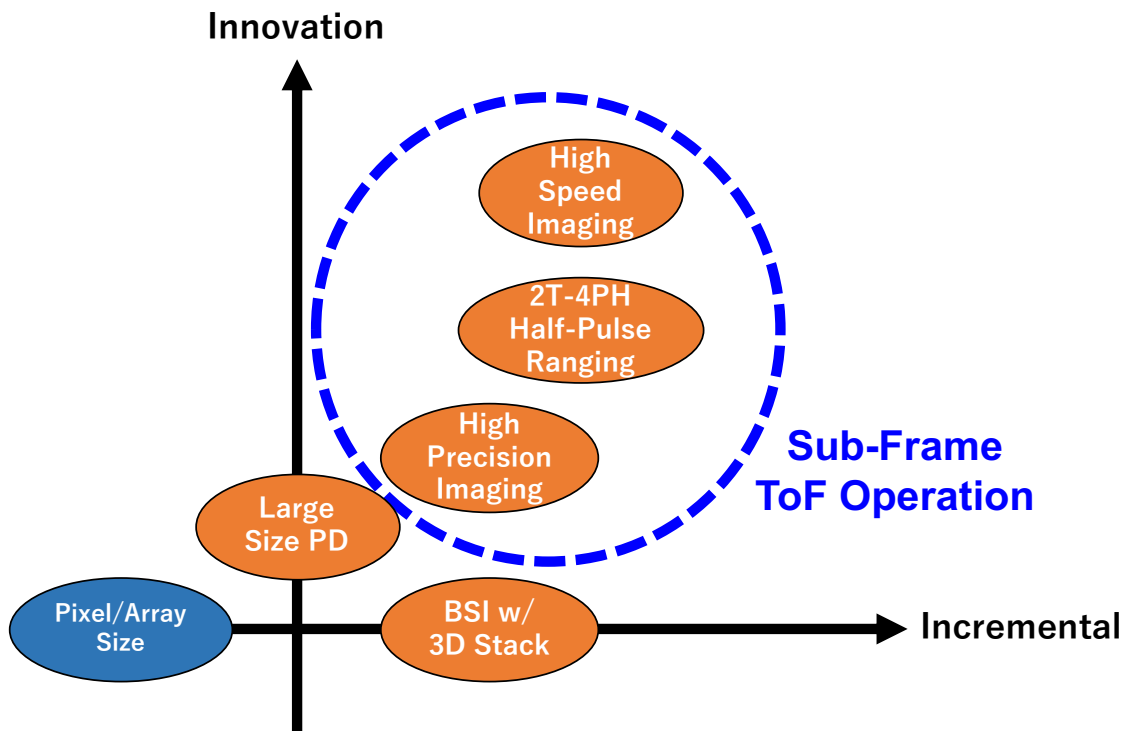
This study developed and demonstrated the unprecedented high-speed range imaging as well as the depth precision enhancement using a prototype iToF range imager. The directions for future improvements were also proposed. The development opens a new avenue for machine vision and show promising potential to enhance the performance and reliability for various 3-D imaging applications.

Reference

[1] R. Hartley and A. Zisserman, *Multiple view geometry in computer vision*. Cambridge university press, 2003.

[2] A. Breitbarth, T. Schardt, C. Kind, J. Brinkmann, P.-G. Dittrich, and G. Notni, "Measurement accuracy and dependence on external influences of the iPhone X TrueDepth sensor," in *Photonics and Education in Measurement Science 2019*, SPIE, 2019, pp. 27–33.

[3] H. Seo *et al.*, "A CMOS LiDAR Sensor with Pre-Post Weighted-Histogramming for Sunlight Immunity over 105 klx and SPAD-based Infinite Interference Canceling," in *2021 Symposium on VLSI Circuits*, IEEE, 2021, pp. 1–2.

[4] 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," in *2021 IEEE International Solid-State Circuits Conference (ISSCC)*, IEEE, 2021, pp. 111–113.

[5] E. Manuzzato, A. Tontini, A. Seljak, and M. Perenzoni, "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," in *2022 IEEE International Solid-State Circuits Conference (ISSCC)*, IEEE, 2022, pp. 96–98.

[6] O. Kumagai *et al.*, "A 189× 600 back-illuminated stacked SPAD direct time-of-flight depth sensor for automotive LiDAR systems," in *2021 IEEE International Solid-State Circuits Conference (ISSCC)*, IEEE, 2021, pp. 110–112.

[7] R. Lange and P. Seitz, "Solid-state time-of-flight range camera," *IEEE J Quantum Electron*, vol. 37, no. 3, pp. 390–397, 2001.

[8] 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, doi: 10.1109/TED.2022.3145762.

[9] S. Shimada *et al.*, "A Back Illuminated 6 μm SPAD Pixel Array with High PDE and Timing Jitter Performance," in *2021 IEEE International Electron Devices Meeting (IEDM)*, IEEE, 2021, pp. 20–21.

[10] K. Morimoto *et al.*, "3.2 megapixel 3D-stacked charge focusing SPAD for low-light imaging and depth sensing," in *2021 IEEE International Electron Devices Meeting (IEDM)*, IEEE, 2021, pp. 20–22.

[11] M. S. 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, doi: 10.1109/JSSC.2021.3112405.

[12] 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)*, IEEE, 2020, pp. 31–33.

[13] C. S. Bamji *et al.*, "IMpixel 65nm BSI 320MHz demodulated TOF Image sensor with 3μm global shutter pixels and analog binning," in *2018 IEEE International Solid-State Circuits Conference-(ISSCC)*, IEEE, 2018, pp. 94–96.

[14] 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 *2020 IEEE International Electron Devices Meeting (IEDM)*, IEEE, 2020, pp. 32–33.

[15] Y. Watanabe, "High-speed optical 3D sensing and its applications," *Advanced Optical Technologies*, vol. 5, no. 5–6, pp. 367–376, Dec. 2016, doi: 10.1515/aot-2016-0047.

[16] I. Gyongy *et al.*, "High-speed vision with a 3D-stacked SPAD image sensor," in *Advanced Photon Counting Techniques XV*, SPIE, 2021, p. 1172105.

[17] S. J. Kim, J. D. K. Kim, B. Kang, and K. Lee, "A CMOS image sensor based on unified pixel architecture with time-division multiplexing scheme for color and depth image acquisition," *IEEE J Solid-State Circuits*, vol. 47, no. 11, pp. 2834–2845, 2012, doi: 10.1109/JSSC.2012.2214179.

[18] T. C. Millar, N. Sarhangnejad, N. Katic, K. Kutulakos, and R. Genov, "The Effect of Pinned Photodiode Shape on Time-of-Flight Demodulation Contrast," *IEEE Trans Electron Devices*, vol. 64, no. 5, pp. 2244–2250, May 2017, doi: 10.1109/TED.2017.2677201.

[19] S. Kawahito *et al.*, "CMOS lock-in pixel image sensors with lateral electric field control for time-resolved imaging," in *Proceedings of the 2013 International Image Sensor Workshop (IISW)*, Snowbird, UT USA, 2013.

[20] H. S. Man, T. Takasawa, T. Akahori, K. Yasutomi, K. Kagawa, and S. Kawahito, "A 413×240-pixel sub-centimeter resolution Time-of-Flight CMOS image sensor with in-pixel background canceling using lateral-electric-field charge modulators," *2014 IEEE International Solid-State Circuits Conference Digest of Technical Papers (ISSCC)*, pp. 130–131, 2014.

[21] S. M. 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 Journal of the Electron*

Devices Society, vol. 3, no. 3, pp. 267–275, May 2015, doi: 10.1109/JEDS.2014.2382689.

[22] 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, doi: 10.1109/JSSC.2018.2789403.

[23] J. Jang *et al.*, "An Ultra-low current operating 5-μm Vertical Field Modulator Pixel for in-direct Time of Flight 3D Sensor," in *Proceedings of the 2021 International Image Sensor Workshop (IISW), Online*, 2021.

[24] D. Stoppa, N. Massari, L. Pancheri, M. Malfatti, M. Perenzoni, and L. Gonzo, "A range image sensor based on 10-μm lock-in pixels in 0.18-μm CMOS imaging technology," in *IEEE Journal of Solid-State Circuits*, Jan. 2011, pp. 248–258. doi: 10.1109/JSSC.2010.2085870.

[25] 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, Nov. 2015, doi: 10.1109/JSSC.2014.2364270.

[26] T. Sawada, K. Ito, M. Nakayama, and S. Kawahito, "TOF range image sensor using a range-shift technique," in *SENSORS, 2008 IEEE*, IEEE, 2008, pp. 1390–1393.

[27] K. Hatakeyama *et al.*, "A Hybrid 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, 2023, doi: 10.1109/JSSC.2023.3238031.

[28] 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.

[29] T.-H. Hsu, T. Liao, N.-A. Lee, and C.-C. Hsieh, "A CMOS time-of-flight depth image sensor with in-pixel background light cancellation and phase shifting readout technique," *IEEE J Solid-State Circuits*, vol. 53, no. 10, pp. 2898–2905, 2018.

[30] 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.

[31] K. Yasutomi, Y. Okura, K. Kagawa, and S. Kawahito, "A Sub-100 μm-Range-Resolution Time-of-Flight Range Image Sensor With Three-Tap Lock-In Pixels, Non-Overlapping Gate Clock, and Reference Plane Sampling," *IEEE J Solid-State Circuits*, vol. 54, no. 8, pp. 2291–2303, Aug. 2019, doi: 10.1109/JSSC.2019.2916310.

[32] K. Yasutomi, Y. Morikawa, S. Imanishi, T. Takasawa, K. Kagawa, and S. Kawahito, "A high-resolution time-of-flight range image sensor with a 3-tap lateral electric field charge modulator," *Technology (Singap World Sci)*, vol. 46, no. 1, 2011.

[33] 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, doi: 10.1109/JSSC.2019.2959502.

[34] 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, doi: 10.1109/LSSC.2022.3225722.

[35] K. Miyauchi *et al.*, "Pixel structure with 10 nsec fully charge transfer time for the 20m frame per second burst CMOS image sensor," in *Image Sensors and Imaging Systems 2014*, SPIE, Mar. 2014, p. 902203. doi: 10.1117/12.2042373.

[36] 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 *Proceedings of the Extended Abstracts of the 2018 International Conference on Solid State Devices and Materials, Tokyo, Japan*, 2018, pp. 9–13.

[37] 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.

[38] Y. Tochigi *et al.*, "A global-shutter CMOS image sensor with readout speed of 1-tpixel/s burst and 780-mpixel/s continuous," *IEEE J Solid-State Circuits*, vol. 48, no. 1, pp. 329–338, 2013, doi: 10.1109/JSSC.2012.2219685.

[39] 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)*, IEEE, 2022, pp. 48–49.

[40] Sophie Caranhac, "Driving gate charge coupled device".

[41] C. Tubert, L. Simony, F. Roy, A. Tournier, L. Pinzelli, and P. Magnan, "High speed dual port pinned-photodiode for time-of-flight imaging," *Proc. IISW*, pp. 1–3, 2009.

[42] 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 (Switzerland)*, vol. 20, no. 1, Jan. 2020, doi: 10.3390/s20010116.

[43] M. Reynolds, J. Doboš, L. Peel, T. Weyrich, and G. J. Brostow, "Capturing Time-of-Flight data with confidence," in *CVPR 2011*, 2011, pp. 945–952. doi: 10.1109/CVPR.2011.5995550.

Reference

[44] A. Tournier *et al.*, "Pixel-to-pixel isolation by deep trench technology: application to CMOS image sensor," in *Proc. Int. image sensor workshop*, 2011, pp. 12–15.

[45] J. Park *et al.*, "Pixel technology for improving IR quantum efficiency of backside-illuminated CMOS image sensor," *International Image Sensor Society*, vol. 4, no. R14, pp. 1–4, 2019.

[46] 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 *2017 IEEE International Electron Devices Meeting (IEDM)*, 2017, pp. 16.4.1-16.4.4. doi: 10.1109/IEDM.2017.8268403.

[47] S. Yokogawa *et al.*, "IR sensitivity enhancement of CMOS Image Sensor with diffractive light trapping pixels," *Sci Rep*, vol. 7, no. 1, pp. 1–9, 2017.

[48] K. Saito *et al.*, "High capacitance density highly reliable textured deep trench SiN capacitors toward 3D integration," *Jpn J Appl Phys*, vol. 60, no. SB, p. SBBC06, 2021.

[49] J.-K. Lee *et al.*, "5.5 A 2.1 e− Temporal Noise and− 105dB Parasitic Light Sensitivity Backside-Illuminated 2.3 μm-Pixel Voltage-Domain Global Shutter CMOS Image Sensor Using High-Capacity DRAM Capacitor Technology," in *2020 IEEE International Solid-State Circuits Conference-(ISSCC)*, IEEE, 2020, pp. 102–104.

[50] Y. Oh *et al.*, "A 140 dB Single-Exposure Dynamic-Range CMOS Image Sensor with In-Pixel DRAM Capacitor," in *2022 International Electron Devices Meeting (IEDM)*, IEEE, 2022, p. 37.

Publication Notes (1ˢᵗ Author)

Journal papers with review

1. **<u>Chia-Chi Kuo</u>** and Rihito Kuroda, "A 4-Tap CMOS Indirect Time-of-Flight Range Imager using In-pixel Memory Array with 10Kfps High-Speed Mode and High Precision Mode," IEEE Journal Solid-State Circuits, June. 2023, doi: 10.1109/JSSC.2023.3281610.

2. **<u>Chia-Chi Kuo</u>** and Rihito Kuroda, "A 2-Tap 4-Phase Indirect Time-of-Flight Ranging Method using Half-Pulse Modulation for Depth Precision Enhancement and Sub-Frame Operation for Motion Artifact Suppression," ITE Transactions on Media Technology and Applications, 2023, Volume 11, Issue 3, Pages 123-129.

Conference proceedings with review

1. **<u>Kuo Chia-Chi</u>** and Rihito 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), Honolulu, HI, USA, 2022, pp. 48–49.

2. **<u>Kuo Chia-Chi</u>** and Rihito Kuroda, " A Half-Pulse 2-Tap Indirect Time-of-Flight Ranging Method with Sub-Frame Operation for Depth Precision Enhancement and Motion Artifact Suppression," in 2023 International Image Sensor Workshop (IISW) Proceedings, Scotland, UK, 2023

Conference presentations without review

1. <u>郭 家祺</u>, 黒田 理人, "画素内アナログメモリアレイを有する 4-Tap Time-of-Flight CMOS イメージセンサによる高精度・高速距離イメージング," 映像情報メディア学会技術報告, Vol. 46, No. 29, pp. 9-12, 2022

2. **Kuo Chia-Chi** and Rihito Kuroda, "A Study on High Precision and High-Speed Time-of-Flight CMOS Image Sensor for Range Imaging," in 34th Microelectronics Conference Proceedings, 2022

Invited talk and Seminar talk

1. 【招待講演】 **Kuo Chia-Chi** and Rihito 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," 2022 年 VLSI シンポジウム国内報告会, July, 2022

2. 【招待講演】 **Kuo Chia-Chi** and Rihito Kuroda, "A High-Speed and High Precision 4-Tap Time-of-Flight CMOS Image Sensor with In-pixel Analog Memory Array for Range Imaging," in KIBME/ITE JOINT SESSION PROGRAM, Nov, 2022

Awards

第 21 回 IEEE EDS Japan Chapter Student Award, Feb. 2023.