IEEE Open Journal of the
Solid-State Circuits Society

Received 25 November 2025; revised 9 January 2026; accepted 25 January 2026. Date of publication 3 February 2026; date of current version 2 March 2026.

Digital Object Identifier 10.1109/OJSSCS.2026.3660622

Ultrafast Time-Compressive CMOS Image Sensors Based on Multitap Charge Modulators for Filming Light-In Flight

KEIICHIRO KAGAWA[1] (Senior Member, IEEE), DAISUKE HAYASHI[2], ARASHI TAKAKURA[3], YUTO UMEKI[2], MICHITAKA YOSHIDA[4], KEITA YASUTOMI[1] (Member, IEEE), SHOJI KAWAHITO[1] (Fellow, IEEE), YOUNGCHEOL CHAE[5] (Senior Member, IEEE), AND HAJIME NAGAHARA[6] (Member, IEEE)

(Invited Paper)

[1]Research Institute of Electronics, Shizuoka University, Hamamatsu 432-8011, Japan
[2]Graduate School of Integrated Science and Technology, Shizuoka University, Hamamatsu 432-8011, Japan
[3]Faculty of Engineering, Shizuoka University, Hamamatsu 432-8011, Japan
[4]Faculty of Environmental, Life, Natural Science and Technology, Okayama University, Okayama 700-8530, Japan
[5]Department of Electrical and Electronic Engineering, Yonsei University, Seoul 120749, South Korea
[6]D3 Center, The University of Osaka, Suita 565-0871, Japan

CORRESPONDING AUTHOR: K. KAGAWA (e-mail: kagawa@idl.rie.shizuoka.ac.jp)

This work was supported in part by JST, CREST, under Grant JPMJCR22C1; and in part by the Grants-in-Aid for Scientific Research (A) under Grant 24H00313.

ABSTRACT Ultrafast time-compressive CMOS image sensors based on multitap charge modulators can capture light-in flight using coded exposure masks on the focal plane. Transient images can then be reconstructed using iterative methods or deep learning models. Although the image sensor is based on indirect time-of-flight (ToF) image sensors, the reconstructed images are equivalent to those captured by direct ToF (D-ToF) image sensors. Important design parameters of the image sensor include the pixel block size and the number of taps of the charge modulator. Several constraints regarding the charge transfer of the multitap charge modulator, the hamming distance between exposure codes at adjacent timings, and the minimal time window duration must be considered when designing exposure codes. The influence of these factors on the fidelity of the reconstructed images is analyzed numerically. The results show that a pixel block size of 4×4 is optimal and that four or more taps are required for light detection and ranging (LiDAR) applications when 32 transient images of light-in flight are reconstructed. To demonstrate LiDAR in a scene with multipath interference, two objects were observed through a weakly diffusive sheet. The temporal resolution, as defined by the clock period of the exposure codes, was 1.65 ns. Multiple reflections were reconstructed using an iterative method (TVAL3) and a deep learning model (ADMM-Net). Although the waveforms of optical pulses reconstructed by TVAL3 are distorted, the amplitudes are more accurate. Conversely, although ADMM-Net reconstructs sharper optical pulses, the amplitudes are inaccurate. To achieve the shorter temporal resolution required for time-resolved diffuse optical tomography (DOT) and fluorescence lifetime imaging (FLIm), the feasibility of heterodyne compression was demonstrated through simulation.

INDEX TERMS CMOS image sensor, compressive imaging, computational photography (CP), multitap charge modulator, transient imaging.

I. INTRODUCTION

TRANSIENT imaging of light-in flight has a variety of applications of quantitative imaging applications, such as light detection and ranging (LiDAR) [1], [2], [3], time-resolved diffuse optical tomography (DOT) [4], and fluorescence lifetime imaging (FLIm) [5]. In these applications, a camera emits short optical pulses with a duration ranging from subnanoseconds to nanoseconds and captures transient images of reflected or fluorescent light from a scene or a sample as their impulse response. This impulse response

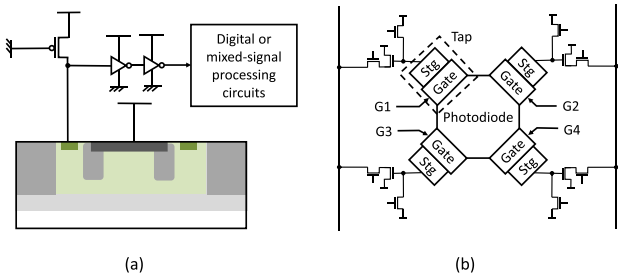


FIGURE 1. Two types of time-resolved photodetectors (a) SPAD and (b) charge modulator. Stg: storage.

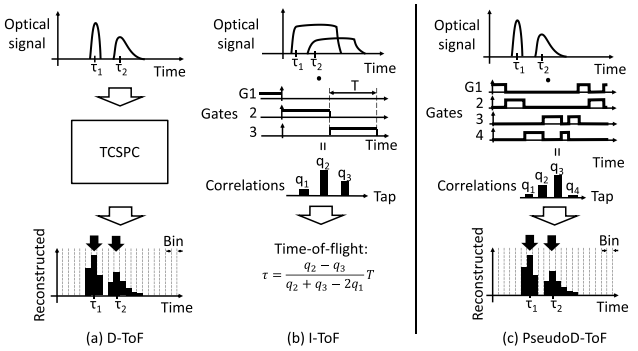


FIGURE 2. Implementations of LiDAR. (a) D-ToF, (b) PM I-ToF, and (c) our pseudoD-ToF that uses a PM I-ToF image sensor but can reconstruct waveforms equivalent to those obtained by D-ToF image sensors.

contains plenty of information. In LiDAR, the optical impulse is only shifted by time of flight (ToF) in a simple scene. ToF is proportional to the target distance or depth. However, multipath interference caused by scattering and multiple reflections can widen and distort the pulse, or increase the number of reflected pulses at different distances [6], [7], [8]. In such cases, acquiring detailed waveforms is crucial for accurately estimating depths. DOT estimates 3-D optical parameter maps of absorption and reduced scattering coefficients from the waveforms of the measured reflected light. The delay of the reflected light depends on the separation between the light source and detector, as well as the optical parameters. The time range of the optical signal in DOT is typically less than a few nanoseconds. FLIm measures the exponentially decaying fluorescence light of fluorophores, which is influenced by their microenvironments, e.g., temperature, pH, and ion concentration. The fluorescence lifetimes of endogenous fluorophores excited by light in the range from near-ultraviolet to blue are typically around one to several nanoseconds. Therefore, the measured time range is from tens to a hundred nanoseconds.

For ultrafast time-resolved imaging, two types of photodetector are used: 1) the single-photon avalanche diode (SPAD) [9], [10], [11], [12], [13], [14], [15], [16], [17], [18], [19]; and 2) the multitap charge modulator [20], [21], [22], [23], [24], [25], [26], [27], [28], [29], [30], [31], [32], [33], [34], [35] (Fig. 1). SPADs convert a single incident photon into a digital pulse [Fig. 1(a)]. This enables high photosensitivity and high temporal resolution to be achieved simultaneously. The charge modulator resembles a basic CMOS image sensor pixel using a pinned photodiode [or a photogate (PG)] and a floating diffusion amplifier [36], [37]. This converts incident photons to electrons, which are transferred to one of the charge storage memories by applying time-window functions G1–G4 to the gates [Fig. 1(b)]. A pair of a gate and a storage memory is referred to as a tap. Many types of charge modulator have been developed, including those based on PGs [20], [21], [22], transfer gates [23], [24], [25], [26], [27], current-assisted photonic demodulation [28], [29], and lateral electric field charge modulation (LFEM) [30], [31], [32], [33], [34], [35]. Note that the charge modulator performs low-level signal processing, namely, correlation between a time-window function and a temporal input optical signal, in the charge domain. Therefore, no processing circuit is required in the pixel, leading to small pixels or a high pixel count.

LiDAR image sensors are generally categorized as either direct ToF (D-ToF) or indirect ToF (I-ToF). In D-ToF, short optical pulses are emitted, and SPADs detect their arrival time. Incident light waveforms are then reconstructed using time-correlated single-photon counting (TCSPC) [Fig. 2(a)] [38]. The bin size of the histogram defines the time resolution. Interpolation in peak position detection can further improve peak position accuracy [14]. One drawback of D-ToF is the circuit area required, particularly for multibit memory. For example, a 10-bit memory composed of D-latches or D-flip-flops consumes a significant amount of area in each pixel. Therefore, fully pixel-parallel TCSPC with many bins is challenging.

On the other hand, I-ToF does not measure the optical waveform directly, but the correlation between the incident light waveform and the time-window functions applied to the gates (G1–G3) [see Fig. 2(b)]. The ToF is then calculated using the detected pixel values. I-ToF can be further categorized as the amplitude continuous wave (AMCW) method or the pulse modulation (PM) method. This article only considers the PM I-ToF due to its similarity to D-ToF.

Charge modulators are well-suited to I-ToF image sensors, as the correlation between the optical signal, or photocurrent signal, and the gate signal can be performed without any in-pixel processing circuits, as mentioned above. Therefore, the pixel count of I-TOF image sensors is much higher than that of D-ToF image sensors. However, the optical waveform detail is lost during measurement because the time-window functions for the gates in I-ToF are much wider than the bin width in D-ToF, and the number of gates is limited (typically, 2–4). To achieve transient imaging with I-ToF image sensors, additional scanning is required [6], [7], [39], which is not suitable for observing dynamic scenes.

If you need to read out the arrival time of every single photon to the off-chip with an extremely low latency, SPADs are the only solution. However, in most imaging applications such as LiDAR, DOT, and FLIm, measurement accuracy and

precision are limited by photon shot noise, and nanosecond or microsecond latency is not required. If more than, e.g., a hundred photons or photogenerated electrons are necessary and a latency of more than a millisecond is acceptable, the benefits of a multitap charge modulator, namely, small charge memory size, high photon rate tolerance, and lower operating voltage, are significantly useful in terms of fabrication cost and device reliability.

The capability of charge modulators to perform low-level correlation operations is closely linked to computational photography (CP). In CP, images are captured after modulation or coding, and more information is reconstructed than in conventional photography, such as depth, wavelength, and light field. Compressive sensing (CS) [40], [41], [42] is one of the most important signal compression and reconstruction schemes in CP. In CS, an optical input signal is multiplied by random binary codes in the optical domain or on the image sensor, and the result is summed up to obtain the measured signal. This process is equivalent to the inner product or correlation between the optical signal and the code. When the dimension or total number of measured data points is smaller than that of the input signal, the reconstruction of the input signals becomes ill-posed. However, if the input signal is sparse, the original signal can be reconstructed from the measured signal. Since charge modulators can perform the necessary correlation between the optical signal and the applied code at each pixel, specialized CMOS image sensors for CS have been developed. One typical application is compressive video [43], [44], [45]. In [45], multiple optical images are multiplied pixelwise with multiple spatially coded random masks using a one-tap or two-tap charge modulator so that one compressed image is read out.

As the signal reconstruction process for CS relied on iterative optimization methods, it was time-consuming and not suitable for real-time processing. However, thanks to significant recent progress in deep learning, real-time reconstruction of compressive video has become possible [46], and the accuracy of the reconstructed images has significantly improved [47], [48].

We have developed ultrafast time-compressive CMOS image sensors for transient imaging with high temporal resolution and pixel count. These sensors are based on the multitap charge modulator and perform focal-plane temporal CS [49], [50], [51], [52], [53], [54], [55]. To reduce circuit complexity, the same exposure codes are applied repeatedly in space. A group of pixels, to which a repetition unit of the exposure codes is applied, is referred to as a pixel block. Thus, the pixel array is composed of pixel blocks, all of which are subjected to the same exposure codes. While the function of these image sensors is similar to that of compressive video image sensors, the time scale is completely different. The time resolution is in the nanosecond regime or shorter, realized by an ultrafast correlation operation with a multitap charge modulator. Fig. 2(c) shows our compressive method, which we call pseudo D-ToF. These image sensors are based on I-ToF image sensor technology. However, they enable us to reconstruct the original optical waveforms, which is the most important feature of D-ToF image sensors. Unlike conventional I-ToF, our method uses random exposure codes, although regular periodic time windows are used in conventional I-ToF. Our CS-based image sensors benefit from the advantages of both I-ToF and compressive video sensors. Therefore, a small pixel size, a high pixel count, and real-time signal reconstruction are possible. The spatiotemporal sampling efficiency of the compressive method is compared with D-ToF image sensors in [53]. Furthermore, as image acquisition is conducted in a single shot, no motion artifacts occur. Consequently, our time-compressive image sensors are ideal for observing dynamic scenes.

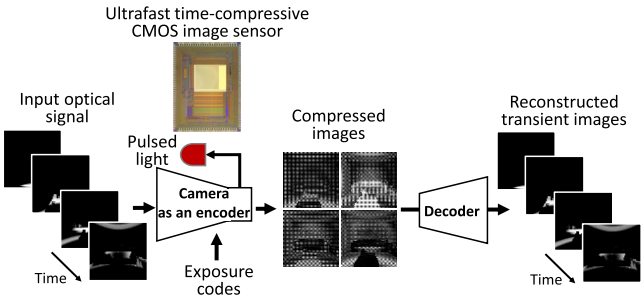


FIGURE 3. Image acquisition and reconstruction.

In this article, the performance of ultrafast time-compressive CMOS image sensors for transient imaging is numerically evaluated for several coding schemes to clarify how to determine the pixel block size and how the coding scheme influences the reconstruction quality. In Section II, the architecture and fundamental operation of a time-compressive CMOS image sensor for transient imaging are explained. The operational constraints of these sensors are mentioned. In Section III, CS, signal reconstruction, and exposure code generation methods are briefly explained. In Section IV, how to determine the two important design parameters, namely, pixel block size and the number of taps, is discussed in terms of reconstruction error. Section V provides demonstrations of compressive LiDAR and compressive FLIm by experiments or simulations. For FLIm, heterodyne compression is introduced to improve temporal resolution.

II. CHARGE-DOMAIN TIME-COMPRESSIVE CMOS IMAGE SENSOR

A. OVERALL SYSTEM AND TRANSIENT IMAGE COMPRESSION

Fig. 3 illustrates the process of image acquisition and reconstruction. The ultrafast time-compressive CMOS image sensor, which is based on a multitap charge modulator, detects and compresses light-in flight. The compressed images, in a number equal to the number of taps, are then read out. Finally, the transient images are reconstructed by a decoder. In this

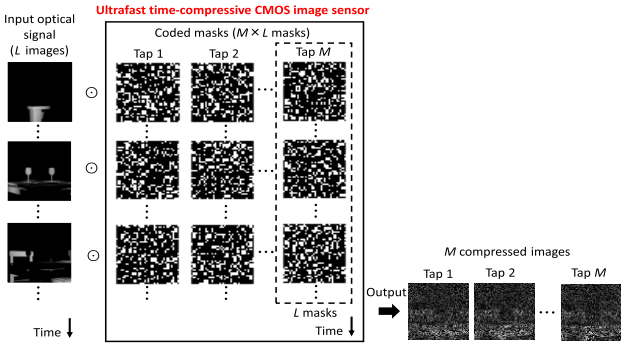


FIGURE 4. Multitap charge modulator-based focal-plane image compression. The number of taps is four.

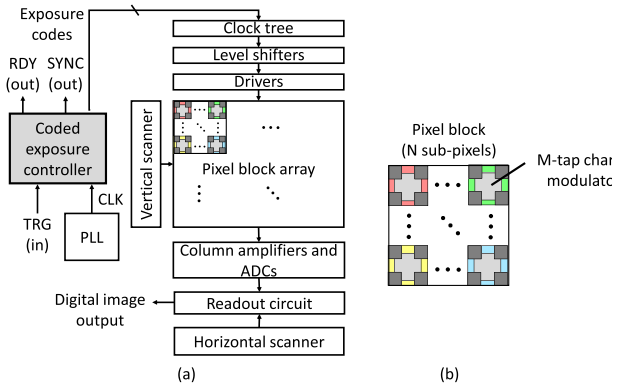


FIGURE 5. (a) Image sensor architecture of ultrafast time-compressive CMOS image sensor. PLL: phase locked loop. ADC: analog to digital converter. (b) Pixel block structure.

process, the image sensor acts as an encoder. Iterative methods or deep learning-based algorithms are used as a decoder.

In compressive video, L binary random-coded masks are prepared to compress the optical signal. The optical distribution on the image sensor at a given moment is multiplied by a coded mask in a pixel-wise operation. The coded mask is then sequentially changed until all L masks have been applied, and the pixelwise products are summed up over an exposure time to generate a pixel value. Finally, the compressed image is read out. This single-tap method is extended to multiple taps to compress more images or to improve the fidelity of reconstructed images in the ultrafast time-compressive CMOS image sensors. As depicted in Fig. 4, multiple sets of coded masks are used. Here, M sets of coded masks (i.e., M taps) are assumed so that M compressed images can be read out.

B. SENSOR ARCHITECTURE

The ultrafast time-compressive CMOS image sensor is similar to conventional I-ToF image sensors in most respects, except for the block structure of the pixel array and the coded exposure controller (see Fig. 5). The image sensor operates in global shutter mode. Therefore, the exposure and image readout periods are separate. The pixel array comprises an array of pixel blocks, each of which is composed of N subpixels. Since

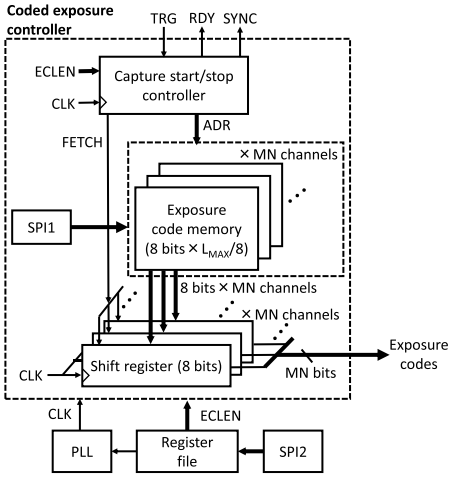


FIGURE 6. Block diagram of coded exposure controller.

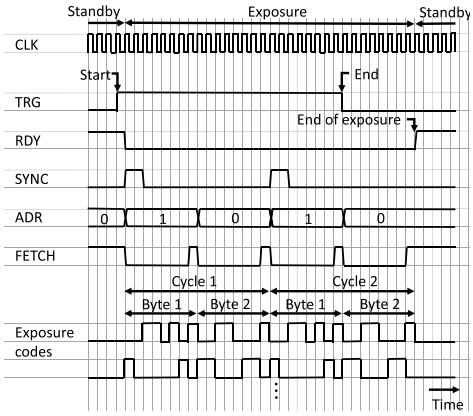


FIGURE 7. Timing chart for ultrafast time-compressive CMOS image sensor.

the pixel uses an M-tap charge modulator, $N \times M$ exposure codes are applied to each block simultaneously during the exposure period. For example, when $N = 2 \times 2$ and $M = 4$, 16 exposure codes are available, and these are applied to all blocks.

Figs. 6 and 7 show a simplified block diagram of the coded exposure controller and a timing chart, respectively. The coded exposure controller is equipped with exposure code memories. The maximum number of coded masks for each tap is denoted by L_{MAX}. As shown in the figures, the coded masks are stored for each tap so that $M \times N$ channels of exposure code memories are prepared. For each channel, the exposure code is stored in bytes. The register ECLEN defines the number of coded masks in bytes and is configured via the serial-to-parallel interface 2 (SPI2). If L-coded masks are applied, ECLEN is set to $L/8$-1 (if ECLEN is 0, eight masks are applied). Exposure codes are written to the exposure code memory via SPI1. The capture start/stop controller controls exposure. In the standby state, the capture start trigger (TRG) and the capture ready signal (RDY) are low and high, respectively. Once TRG turns high, RDY turns low, and one byte of the exposure codes is read out to the shift registers at a

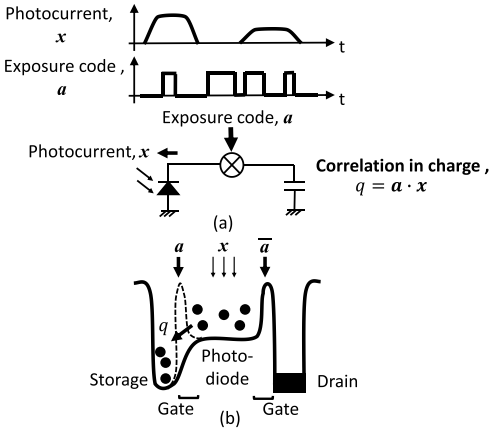


FIGURE 8. One-tap charge modulator with a drain. (a) Schematic and (b) potential diagram.

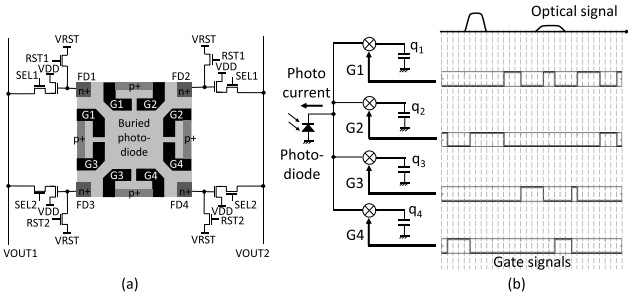


FIGURE 9. Four-tap charge modulator without a drain.(a) LEFM-based pixel structure and (b) schematic. FD: floating diffusion. VRST: reset voltage. VDD: power supply voltage. RST1-2: reset signals. SEL1-2: select signals. VOUT1-2: output signals.

time in parallel at each channel, controlled by the code fetch signal (FETCH). The memory address signal (ADR) counts up from 0 to (L/8)-1. Then, the codes in the shift registers are read out serially for all $M \times N$ channels in parallel. This architecture is ideal for high-speed, low-power operation because the maximum operating speed is largely determined by the simple shift registers. In one cycle, L-coded masks are read out. If TRG is high at the end of the cycle, the next cycle begins. Otherwise, exposure ends, and RDY returns to high, indicating the standby state. Then, the captured images are read out. A SYNC signal is output at the beginning of each cycle and is used to TRG the emission of pulsed light.

C. MULTITAP CHARGE MODULATORS

The principles and implementations of multitap charge modulators are summarized briefly here. Fig. 8 shows the schematic and potential diagram of a one-tap charge modulator with a drain that can be considered a two-tap charge modulator with one output. As shown in Fig. 8(a), the temporal photocurrent signal $\boldsymbol{x}$ is generated from an incident optical signal. The amount of charges q flowing into the capacitor is given by the correlation between the photocurrent $\boldsymbol{x}$ and the temporal exposure code $\boldsymbol{a}$, which gates the photocurrent. Core processing of correlation involves multiplying two functions, $\boldsymbol{x}$ and $\boldsymbol{a}$, and integrating them, which is performed by the charge modulator. Note that the generation of $\boldsymbol{a}$ and the signal readout are conducted by peripheral circuits outside the pixel array.

This function can be embodied by a photonic device, the potential diagram of which is depicted in Fig. 8(b). A storage memory and a drain are placed beside a photodiode. The potential heights between the photodiode and the storage memory (or the drain), referred to as gates, are modulated according to the exposure code $\boldsymbol{a}$, which is typically binary because accurate analog modulation is challenging due to the pixelwise variation of gates' potential heights. The photocurrent blocked by the gate in Fig. 8(a) is transferred to the drain, which is connected to a constant voltage. The inversion of the exposure code is applied to the drain gate.

Each tap consists of a charge memory and a gate. Preparing multiple taps for a photodiode realizes a multitap charge modulator. Fig. 9(a) shows an example of a four-tap charge modulator based on LEFM. The gate signals (G1–G4) control the charge flow. Only one tap can be activated at a time in the multitap charge modulator to avoid uncertain charge transfer to multiple taps, as shown in Fig. 9(b). There are options for the charge storage device and the drain. The charge modulator in Fig. 9(a) has no dedicated drain; it stores charges in the floating diffusion. To perform true correlated double sampling (CDS) to eliminate reset noise and to share the read and reset circuits to reduce pixel size, a pinned storage diode or storage gate is used for charge storage [21], [22], [25], [33], [34]. If the pixel has no drain, charges generated during the image readout period will overflow and contaminate signals in storage memories. For PM I-ToF image sensors, multitap charge modulators with a dedicated drain have been developed [26], [27], [30], [31], [32], [33], [34], [35], which are also effective at suppressing ambient light. To extend the measurable range, six-tap and eight-tap charge modulators have been developed [35], [56], [57].

D. CONSTRAINTS ON EXPOSURE CODES

Due to the electrical and device limitations of the ultrafast time-compressive CMOS image sensor, several constraints must be considered when designing exposure codes. Therefore, it is not always possible to apply arbitrary or theoretically optimal codes.

1) *Tap Constraint:* Only one tap, including the drain, can be activated in each pixel.
2) *Hamming Distance Constraint:* The maximum hamming distance between the exposure codes at adjacent timings is limited.
3) *Time-Window Constraint:* The minimal time window duration for both on and off is limited.

These constraints are caused by: 1) multitap charge modulators; 2) the IR drop of the drivers' power supply lines; and 3) the speed of the drivers and modulators, respectively. The influences of these constraints will be discussed via simulation in Section IV.

III. IMAGE ACQUISITION AND RECONSTRUCTION BASED ON CS

A. COMPRESSIVE SENSING

Because optical signals received by the image sensor in the short pulse-based LiDAR are very sparse, we can compress the signal in the physical domain based on the CS. The mathematical representation of CS is summarized below. When we consider an α-dimensional column vector $\boldsymbol{x}$, an $\beta \times \alpha$ matrix $\boldsymbol{A}$, and an β-dimensional column vector $\boldsymbol{y}$, the relationship is denoted by

$$\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}. \tag{1}$$

Note that $\boldsymbol{x}$ is the original temporal optical signal, and $\boldsymbol{y}$ is the signal measured through a known measurement matrix $\boldsymbol{A}$. When $\alpha > \beta$, the original signal $\boldsymbol{x}$ is compressed. We can reconstruct the original optical signal $\boldsymbol{x}$ by solving an inverse problem from $\boldsymbol{y}$ and $\boldsymbol{A}$ based on a sparsity constraint. Sparsity means that most elements of $\boldsymbol{x}$ or $\Phi\boldsymbol{x}$ (Φ is any $\alpha \times \alpha$ transformation matrix) are zero, and only a limited number of the elements have nonzero values. $\boldsymbol{A}$ describes the spatial point spread function of imaging optics, exposure codes, and temporal responses of the image sensor and the light source.

B. FORWARD MODEL OF IMAGE ACQUISITION

The signal compression performed by the spatiotemporal compressive CMOS image sensor can be formulated as follows [55]:

$$\begin{aligned} &p\left(X, Y, i\right) \\ &= \iint_{S(X,Y)} \int \left\{SR\left(x, y, t\right) * l\left(t\right) * PSF\left(x, y\right) * h_{PD}\left(t\right)\right\} \\ &\quad \star \left\{h_i\left(t\right) * g_{X,Y,i}\left(t\right)\right\} dtdxdy. \end{aligned} \tag{2}$$

The operators $*$ and $\star$ represent convolution and cross correlation, respectively. X and Y are the subpixel indices, and i is the tap index. $S(X, Y)$ is the area of the pixel (X, Y). $SR(x, y, t)$ is the reflected optical signal on the image sensor and corresponds to the scene's impulse response to be captured. $p(X, Y, i)$ is the pixel value for tap i of the pixel (X, Y). Temporal light source waveform, $l(t)$, a spatial point spread function of imaging optics, $PSF(x, y)$, the response of a photodiode, $h_{PD}(t)$, an exposure code at tap i of the pixel (X, Y), $g_{X,Y,i}(t)$, and the temporal response of tap i, $h_i(t)$ are considered in (2). The overall flow is shown in Fig. 10, where $h_i(t)$ and $h_{PD}(t)$ are summarized as $h_i(t)$.

Equation (2) is transformed as follows:

$$\begin{aligned} &p\left(X, Y, i\right) \\ &= \iint_{S(X,Y)} \int SR\left(x, y, t\right) \cdot IRF_{X,Y,i}\left(t\right) dtdxdy \end{aligned} \tag{3}$$

$$\begin{aligned} &IRF_{X,Y,i}\left(t\right) \\ &= \left\{g_{X,Y,i}\left(t\right) * h_i\left(t\right)\right\} \star \left\{l\left(t\right) * PSF\left(x, y\right) * h_{PD}\left(t\right)\right\} \end{aligned} \tag{4}$$

where $IRF_{X,Y,i}(t)$ is the instrumental response function at the tap of the pixel (X, Y) for a given exposure code. $IRF_{X,Y,i}(t)$ is described further in (4)

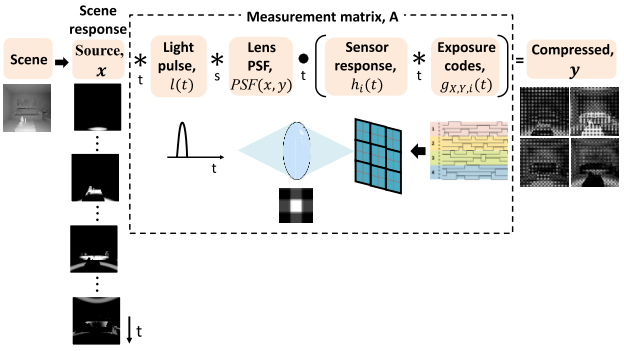


FIGURE 10. Forward model of ultrafast time-compressive CMOS image sensor. "s" and "t" under the operators mean spatial and temporal, respectively. ● is a correlation operator, which is shown as an integral product in (2) and (3).

Equations (3) and (4) are represented by a matrix $\boldsymbol{A}$, as shown in (1). The scene's response $\boldsymbol{x}$ is compressed to $\boldsymbol{y}$ by the image sensor through the measurement matrix $\boldsymbol{A}$, where $\boldsymbol{x}$, $\boldsymbol{y}$, and $\boldsymbol{A}$ correspond to $SR(x, y, t)$, $p(X, Y, i)$, and $IRF_{X,Y,i}(t)$, respectively.

C. SIGNAL RECONSTRUCTION METHODS

1) ITERATIVE METHODS WITH REGULARIZATION

There are several options for signal reconstruction. Conventional methods are based on iterative methods. Since CS is an ill-posed problem, regularization using the L1 norm or total variation (TV) is often employed. Robust and efficient methods have been proposed [58], [59]. However, since reconstruction requires many iterations, these methods are not suitable for real-time processing.

2) DEEP LEARNING

Another option is a deep learning approach. For example, ADMM-Net [60], [61] is a promising candidate due to its short and constant reconstruction time. However, ADMM-Net requires a lot of memory due to its multistage structure and numerous network parameters. To reduce the network size and required memory, we confirmed the effectiveness of combining a normalized transposed observation matrix and a U-Net [54]. The compressed signal is partially reconstructed by multiplying the transposed measurement matrix by the captured compressed images. Then, the U-Net estimates only a depth map, while depth errors by MPI are corrected simultaneously. The memory size and processing time decreased to approximately one-tenth. Although the processing time is longer than that of the aforementioned methods, signal reconstruction methods based on diffusion models [47] and deep image priors [48] have achieved accurate image reconstruction for solving general inverse problems.

D. DESIGNING EXPOSURE CODES

1) RANDOM GENERATION AND SELECTION

The simplest yet effective method for designing exposure codes is random generation and selection [50]. A dataset of transient images is used to verify the fidelity of the reconstructed transient images with given exposure codes. Many

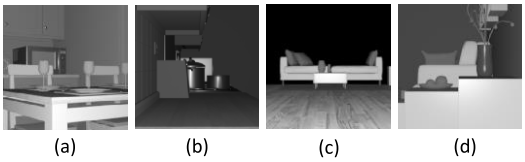


FIGURE 11. Examples of transient images for performance evaluation. Four out of six intensity images are shown. (a) and (b) Kitchen from different angles. (c) and (d) Living-room-3 from different angles.

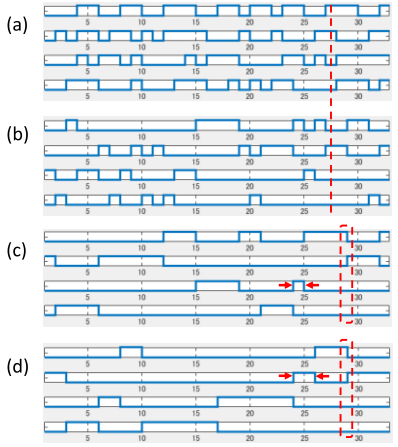


FIGURE 12. Examples of exposure codes with (a) no constraint, (b) tap constraint, (c) Hamming distance constraint, and (d) time-window constraint.

sets of exposure codes are randomly generated while satisfying the design constraints mentioned in Section II-D. The set of exposure codes that yields the smallest mean square error (mse) for the dataset is selected.

2) DEEP SENSING

Exposure codes can be optimized while a deep learning model is trained. Deep sensing [62] is one method of optimizing a neural network and the physical parameters of the optics or electronics in an imaging system. In [53] and [54], deep sensing was applied to pseudo D-ToF. This optimization method yields optimal exposure codes tailored to the dataset's domain.

IV. PERFORMANCE COMPARISON

A. SIMULATION SETTINGS

The pixel block size and the number of charge modulator taps are crucial design parameters of an ultrafast time-compressive CMOS image sensor. These parameters, along with constraints on the exposure code, influence the accuracy of reconstructed transient images. This study numerically analyzes the influence of these factors for LiDAR applications. The dataset was generated by a transient renderer [63] for 3-D models [64]. Reconstruction fidelity is evaluated using mse for six sets of transient images. Four of these sets are shown in Fig. 11.

The constraints explained in Section II-D are considered in the simulation. Example exposure codes are shown in Fig. 12.

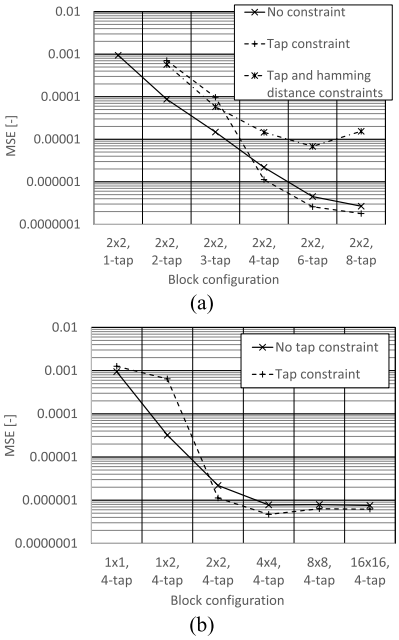


FIGURE 13. MSE of the reconstructed transient images. (a) Number of taps is changed from 1 to 8 for 2 × 2 pixels per block. (b) Block size is changed from 1 × 1 to 16 × 16.

When there are no constraints, the exposure code for each channel is generated completely randomly. Therefore, the mean probabilities of 0 (inactivated) and that of 1 (activated) are both 50%. With the tap constraint, one tap of each charge modulator is randomly selected. For M-tap modulators, the mean probability that a tap will be selected at any given moment is $1/M$. For the hamming distance constraint, a hamming distance of 2 is considered. In this case, only two taps in a pixel block change at a time. For the time-window constraint, the minimum time-window duration for 0 and 1 is two clocks.

The following parameters are assumed in the simulation. The pixel count is 128 × 128. The charge modulator has four taps. The pixel block is composed of 2 × 2 pixels. The number of coded masks, equivalent to the number of reconstructed images, is 32. Therefore, the compression ratio is 8 (= 32/4) for these settings. The number of exposure code sets is 20. TVAL3 [58] is used as an inverse problem solver. The parameter settings are as follows: maxit = 200, TVnorm = 1, nonneg = true, isreal = true, TVL2 = true, mu = 2^9, beta = 2^5, tol = 1E-6, and maxcnt = 10. No noise was considered to evaluate the maximum performance.

B. SENSOR PARAMETERS VERSUS PERFORMANCE

Fig. 13 summarizes the fidelity of the reconstructed transient images using mse. Fig. 13(a) shows the impact of the number of charge modulator taps when the pixel block size is set to 2 × 2. For the no constraint and the tap constraint, the mse decreases with the number of taps because the compression ratio decreases. In this simulation, the number of reconstructed transient images is set to 32. If a higher temporal

TABLE 1. Reconstruction errors for exposure codes constraints.

Constraints	MSE [-]
No constraint (completely random)	2.19×10^{-6}
Tap constraint	1.12×10^{-6}
Tap constraint + time-window constraint	3.57×10^{-6}
Tap constraint + hamming distance constraint	1.46×10^{-5}
Tap constraint + hamming distance constraint + time-window constraint	1.37×10^{-5}

resolution is needed without increasing the compression ratio (because larger compression ratios will cause larger reconstruction errors), more taps are required. However, as the number of taps increases, the pixel size also increases. The results show that the hamming distance constraint is harmful and results in a larger mse than the other cases. The tap constraint results in a smaller mse than the no constraint. This is because fewer data points are multiplexed to the pixel value. Note that the mean activation probability of a tap is 0.5 for no constraint and $1/M$ (0.25 for $M = 4$) for the tap constraint.

Selecting the pixel block size is crucial because the number of gate signal lines increases accordingly when utilizing the sensor architecture shown in Fig. 5. For example, with a pixel block size of 2×2, the number of gate signal lines doubles compared to conventional I-ToF image sensors without a pixel block structure. Therefore, a large block size is not suitable for small pixels. Fig. 13(b) shows that the optimal pixel block size is 4×4 for the given dataset. Note that larger pixel blocks do not improve the mse. This is because the transient images of light-in flight are nearly continuous in time and space except at the boundaries of objects. A block size of 2×2 provides additional information to distinguish between neighboring pixels, which is most effective to reduce mse. However, as the block size increases, its effectiveness decreases. To achieve small pixels, a pixel block size of 2×2 and a four-tap charge modulator are good options.

Table 1 discusses the influence of the time-window constraint. Although the time-window constraint increases the mse when only the tap constraint is considered, the effect of the time-window constraint is smaller than that of the Hamming distance constraint. When the tap and hamming distance constraints are applied simultaneously, the addition of the time-window constraint does not significantly change the mse.

As discussed above, the hamming distance constraint is the most harmful. Reducing the resistance of the pixel drivers' power lines and the inductance of bonding wires can alleviate the tap constraint. However, the hamming distance constraint limits the frequency of gate switching, which suppresses power dissipation. Even if the hamming distance constraint is solved, we must consider the tradeoff between the hamming distance of exposure codes and power dissipation.

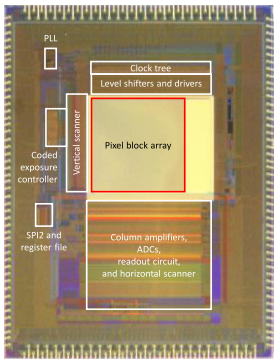


FIGURE 14. Micrograph of a prototype ultrafast time-compressive CMOS image sensor.

TABLE 2. Specifications of a prototype image sensor.

Parameter	Value
Process	0.11μm CMOS image sensor process (FSI)
Number of sub-pixels	188×212
Sub-pixel size	11.2μm×11.2μm
Number of taps per modulator	4 (no drain)
Max. modulation frequency	~300MHz@λ=660nm
Frame rate	21 fps
Max. clock frequency of shutter controller	~800MHz
Number of exposure codes	16
Number of coded masks	8 to 256 bits (by 8 bits)
Column ADC	Folding integration and cyclic

V. DEMONSTRATIONS BY EXPERIMENTS AND SIMULATIONS

A. PROTOTYPE IMAGE SENSOR

Fig. 14 and Table 2 show a micrograph and the specifications of a prototype ultrafast time-compressive CMOS image sensor. A four-tap LEFM without a dedicated drain is utilized as a charge modulator. This sensor can use one of the four taps as a drain, which reduces the number of taps. Therefore, the drain was not used in this demonstration. The measurement was conducted in a dark room. This charge modulator has a high photosensitivity in the visible light region but has a slow operation speed for near-infrared light. Therefore, visible light sources were used for the demonstration.

B. MULTIPATH INTERFERENCE IN LIDAR

To demonstrate pseudo D-ToF, a scene with dual-path interference was captured. The wavelength and the duration of the light pulse were 660 nm and 2.55-ns FWHM, respectively. The frequency of the coded exposure controller was 607 MHz. Therefore, one clock period was 1.65 ns. The exposure time was 55 ms, during which 1 044 736 optical pulses were emitted. Fig. 15(a) and (b) shows a visible-light image of the scene, and images captured by the prototype image sensor, respectively. For reference, a was placed 1.6 m

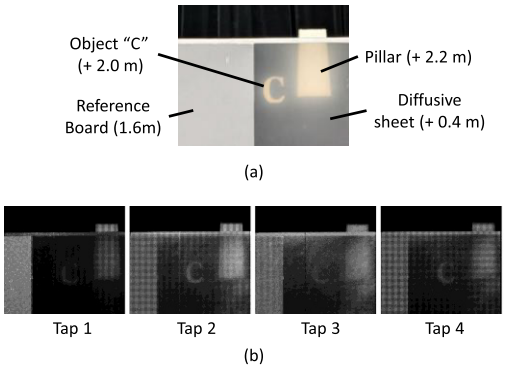


FIGURE 15. Experiments of acquiring compressed images with a prototype CMOS image sensor for a scene with multipath interference.(a) Scene setup. (b) Captured images after averaging 200 images. Relative depths are based on the depth of the reference board.

away from the camera. Two objects, "C" and a pillar, were placed behind a weakly diffusive sheet. Their relative distances from the reference are shown in the figure. The exposure codes were optimized based on deep sensing. 32 coded masks were used for compression. The hamming distance and time-window constraints explained in Section IV-A were considered in the optimization. The signal reconstructions by TVAL3 and ADMM-Net were compared. The parameter settings for ADMM-Net are summarized in Table 3. The parameter settings for TVAL3 are as follows: maxit = 200, TVnorm = 1, nonneg = true, isreal = false, TVL2 = true, mu = 2^{10}, beta = 2^5, tol = 1E-3, and maxcnt = 5. In the reconstruction, the measurement matrix **A** that was oversampled by a factor of four was used to improve the depth accuracy. Fig. 16 shows examples of the measured temporal responses used for the matrix **A** with and without oversampling. Fig. 16(a) shows the temporal response measured at a sampling frequency of 607 MHz, which is equivalent to the operating frequency of the coded exposure controller. Fig. 16(b) was measured with 4× oversampling, where the sampling frequency was 2.428 GHz. While 32 images are reconstructed with normal sampling, 128 images are reconstructed as transient images with the 4× oversampling. To detect the depth, the peak positions are detected first. Then, by using the neighboring data points, are used to refine the depth using parabolic fitting [52], [53]. 4× oversampling reduced the sampling pitch from 1.65 ns (equivalent to a depth of 0.495 m) to 0.4125 ns (0.124 m). The refinement provides a similar effect to the interpolation used in [52]; the depth becomes continuous rather than stepwise.

The processing times of ADMM-Net with 4× oversampling and without oversampling were about 300 and 35 ms, respectively. For processing, a desktop computer with Intel Core i5-12400 and NVIDIA RTX 3060 was used.

Fig. 17 shows the estimated depths in a point-cloud representation. The reference in Fig. 17(a) was measured by scanning the timing of the laser emission for a single time window with a duration of 2 bits (3.3 ns). No reconstruction

TABLE 3. Parameters for ADMM-net.

Parameter	For LiDAR	For FLIm
Number of epochs	1000	100
Loss function	RMSE+MAE	RMSE
Batch size	2	
Learning rate	0.001	
Optimization algorithm	Adam	
Added noise	Shot noise (full well capacity of 20,000e-) Gaussian noise (σ=40e-)	
Dataset	63 (training) 10 (validation) 10 (test)	446 (training) 56 (validation) 56 (test)
Stage count of ADMM-Net	9	
Channel count (CN)	128	32

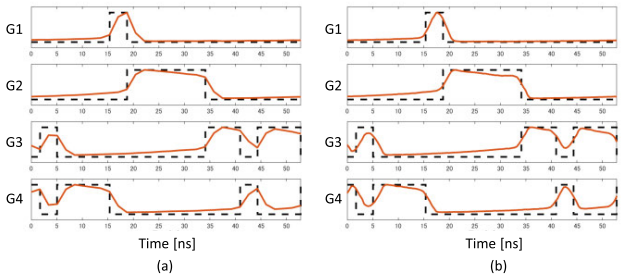


FIGURE 16. Measured temporal response for the subpixel 1 with (a) normal sampling at 607 MHz and with (b) 4× oversampling. Dashed lines show the given binary exposure codes.

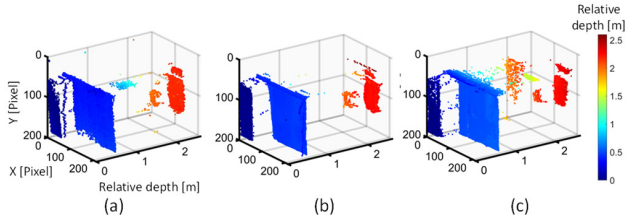


FIGURE 17. Estimated depths by peak detection and parabolic interpolation. (a) Reference measured by scanning a single time window, (b) TVAL3, and (c) ADMM-Net.

processing was conducted for the reference. Although the reference is similar to the ideal scene response, namely, the ground truth, there is a difference: it is convolved with the IRF, which includes the responses of the light source and image sensor. In this experiment, the diffusive sheet introduces multiple reflections. However, this type of MPI does not alter the peak positions of the reflected light. Fig. 18 compares the measured and reconstructed waveforms at points A–C, as shown in Fig. 18(d). At points A and B, two bounces were observed due to weak reflections from the diffusive sheet. At point C, only a single bounce was observed. The waveforms reconstructed by TVAL3 are closer to the reference waveforms (REF) than those by ADMM-Net in terms of peak position and the ratio of the peak amplitudes. However, they became rectangular due to the TV regularization term. The original waveforms are smooth,

TABLE 4. Comparison of estimated depths (averaged for 3 × 3 pixels). Positions A and B are shown in Fig. 18(d).

Positions	Reference [m]	TVAL3 [m]	ADMM-Net [m]
A (1st bounce)	0.394	0.451 (+14.4%)	0.364 (-7.6%)
A (2nd bounce)	2.015	2.020 (+2.5%)	2.131 (+5.8%)
B (1st bounce)	0.389	0.469 (+20.6%)	0.575 (+47.8%)
B (2nd bounce)	2.173	2.245 (+3.3%)	2.280 (+4.9%)
C	2.203	2.221 (+0.8%)	2.356 (+6.9%)

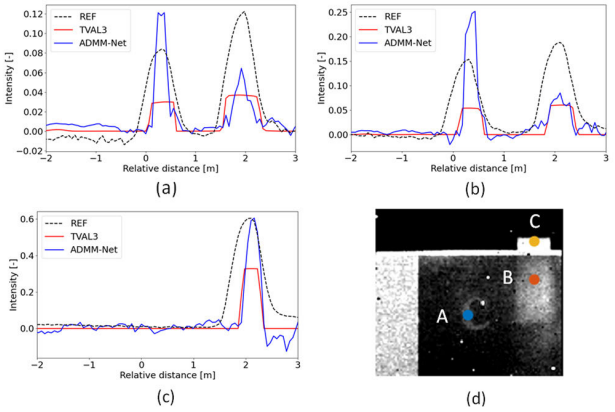


FIGURE 18. Comparison of reconstructed optical waveforms. Waveforms at (a) point A, (b) point B, and (c) point C. Points A–C are shown in (d) intensity image. REF: reference.

as shown by the reference waveforms (REF) in Fig. 18. The ADMM-Net waveforms are sharper than the reference and TVAL3 waveforms due to the deconvolution effect of ADMM-Net. There are discrepancies regarding the amplitude. In addition, as shown in the relative errors in Table 4, the peak positions and the depths differ from those of the reference. These discrepancies may be due to a mismatch between the dataset and the real measured data, as the dataset was generated by simulation. For example, although the directivity of reflectance, surface roughness, and texture differ for each object in real scenes, these factors are not considered in simulations. Building a dataset based on real measurements is crucial to bridging the gap.

The number of detectable reflections is of interest. It is essentially determined by the range and the width of the reconstructed optical pulse waveform. If two adjacent pulses are closely spaced in time and produce only one peak, it is difficult to detect two peak positions. When the minimal time interval between two adjacent reflections, which give two distinctive peaks, is given by ΔT, and the measurable range is T, the maximum number of detectable reflections is given by $T/\Delta T$. Using shorter optical pulses and a higher exposure controller frequency can reduce ΔT to increase the number of detectable pulses. Investigating the relation among ΔT, the number of taps, and compression ratio is important for future work.

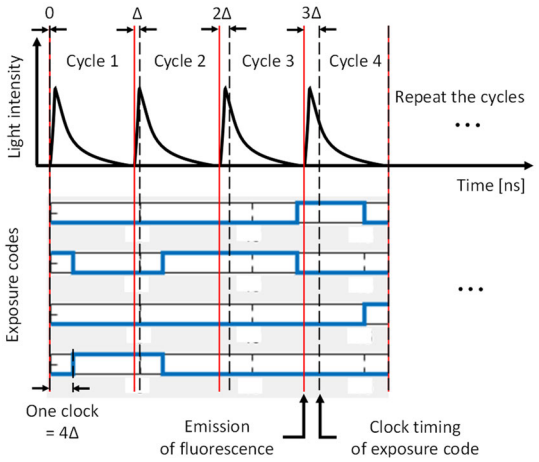


FIGURE 19. Schematic of heterodyne compression. In this case, the unit delay, Δ, is 1/4 of the clock period of the exposure codes.

C. FLIM WITH HETERODYNE COMPRESSION

The time resolution of reconstructed transient images is defined by the clock period of the coded exposure controller during image acquisition. For this prototype image sensor, this period is longer than 1.2 ns. However, FLIm and DOT require time resolutions of less than 1 ns. A heterodyne method [65], [66] can achieve such a short time resolution. As depicted in Fig. 19, the periods of the excitation laser (or fluorescence) emission and the exposure codes differ slightly. In the figure, the fluorescence emission timing is incrementally delayed by Δ (1/4 of one clock period) for each cycle. Therefore, the effective time resolution is Δ. When using ordinary I-ToF-based image sensors for FLIm, many images are captured while scanning the laser's emission timing [33]. Thus, the data acquisition time is longer than with D-ToF-based image sensors. Conversely, heterodyne compression scans the laser emission during the exposure, and the images are captured only once. Therefore, the measurement time is shorter. If the clock cycle is 1.65 ns and the number of scan steps is eight, the temporal sampling pitch, or Δ, becomes 206 ps. This is sufficient for FLIm.

The aforementioned heterodyne compression was verified through simulation. A dataset was created for the simulation using a database of fluorescent images [67]. Each sample in the database is composed of four channels. Here, a channel refers to a specific fluorophore, and each fluorophore is expected to have a different fluorescence lifetime. Three of these channels were selected, and lifetimes (τ_1, τ_2, τ_3) were randomly assigned to each channel in each sample. When generating the dataset, we assumed that τ_1, τ_2, and τ_3 followed uniform distributions. Thus, the lifetimes were randomly determined using uniform distributions with ranges of [0.5, 1.5 ns], [1.5, 2.5 ns], and [2.5, 3.5 ns]. The pixel value for each channel was considered the

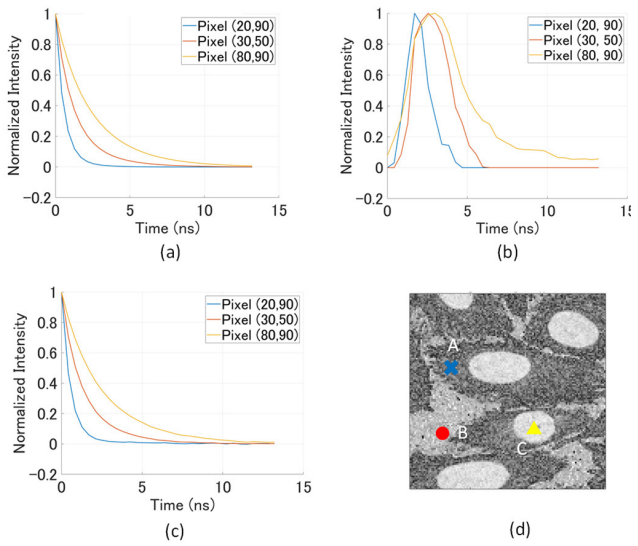


FIGURE 20. Comparison of reconstructed optical waveforms. (a)–(c) Waveforms at points A–C in (d). (a) Ground truth. (b) TVAL3. (c) ADMM-Net. (d) Intensity image of a biological sample in the database.

TABLE 5. Comparison of reconstruction error in FLIm.

Reconstruction algorithm	MSE [-]
TVAL3	0.1433
ADMM-Net	0.0049

weight of the exponentially decaying fluorescence component. The synthesized fluorescence waveform is denoted as follows:

$$I_{X,Y}(t) = \sum_{i=1}^{3} \frac{P_i(X,Y)}{\sum_{j=1}^{3} P_j(X,Y)} \exp\left(-\frac{t}{\tau_i}\right). \quad (5)$$

Here, $P_i(X, Y)$ presents the pixel value of a sample at channel i for pixel (X, Y). The same exposure codes used in Section V-B were applied. The delay step, Δ, was set to 1/8 of a clock period, equivalent to 206 ps at a clock frequency of 607 MHz. The deep learning model settings are summarized in Table 3. The settings for TVAL3 are the same as in Section V-B.

Fig. 20 compares the ground truth with the reconstructed fluorescent waveforms by TVAL3 and ADMM-Net. Although TVAL3 reconstructed the decaying curves, the waveforms differ from an exponentially decaying function. This is because TV regularization, which is used to solve CS problems, tends to reproduce stepwise waveforms by nature. Therefore, it is difficult to accurately reconstruct sharply decaying functions observed in FLIm. In contrast, ADMM-Net accurately reconstructed the waveforms. As shown in Table 5, ADMM-Net outperforms TVAL3 in terms of the fidelity of the reconstructed fluorescent decay.

VI. CONCLUSION

The concept, implementation, performance comparison, and demonstrations of ultrafast time-compressive CMOS image sensors based on multitap charge modulators are discussed. These image sensors are based on I-ToF image sensors. However, after signal reconstruction, the optical waveforms equivalent to those obtained by D-ToF image sensors are acquired. For the signal reconstruction, we can use not only conventional iterative methods with the TV regularization but also deep learning models for accurate and fast reconstruction. Two important design parameters are the pixel block size and the number of taps of the charge modulator. The influence of these parameters on the signal reconstruction fidelity was numerically analyzed based on an iterative signal reconstruction method. The results show that at least four taps are necessary and a pixel block size of 4×4 is optimal for LiDAR applications. Considering the number of gate signal lines, which is related to the pixel pitch, a pixel block size of 2×2 can be a good option. Several constraints should be considered in the design of exposure codes. The hamming distance constraint limits the number of bits that can change at a time, significantly degrading the fidelity of reconstructed transient images. This problem can be alleviated by decreasing the inductance of the bonding wires and the resistance of the pixel driver power supply. However, increasing the hamming distance of exposure codes increases power dissipation. To demonstrate LiDAR in a scene with multipath interference, two objects were observed by a prototype image sensor with four-tap charge modulators and a pixel block size of 2×2 through a weakly diffusive sheet. The temporal resolution, or one clock period for the exposure codes, was 1.65 ns. The exposure code length was 32. Oversampling with a factor of four was used in signal reconstruction. Dual reflections were successfully reconstructed using an iterative method (TVAL3) and a deep learning model (ADMM-Net). However, the waveforms reconstructed by TVAL3 were distorted and became rectangular. The waveforms reconstructed by ADMM-Net appeared more natural and were sharpened due to the deconvolution effect of ADMM-Net. The peak positions were slightly shifted, probably due to a mismatch between the simulated and real measured data. To achieve the shorter time resolution required for time-resolved DOT and FLIm, heterodyne compression was demonstrated through simulation.

REFERENCES

[1] F. Remondino and D. Stoppa, *TOF Range-Imaging Cameras*. Heidelberg, Germany: Springer, 2013.
[2] R. Horaud, M. Hansard, G. Evangelidis, and C. Ménier, "An overview of depth cameras and range scanners based on time-of-flight technologies," *Mach. Vis. Appl.*, vol. 27, no. 7, pp. 1005–1020, Oct. 2016.
[3] R. Lange and P. Seitz, "Solid-state time-of-flight range camera," *IEEE J. Quantum Electron.*, vol. 37, no. 3, pp. 390–397, Mar. 2001.
[4] T. Vo-Dinh, *Biomedical Photonics Handbook*. Boca Raton, FL, USA: CRC Press, 2003, ch. 21.
[5] A. Periasamy and R. M. Clegg, *FLIM Microscopy in Biology and Medicine*. Boca Raton, FL, USA: CRC Press, 2009.
[6] F. Heide, M. Hullin, J. Gregson, and W. Heidrich, "Light-in-flight: Transient imaging using photonic mixer devices," in *Proc. ACM SIGGRAPH Emerg. Technol.*, Jul. 2013, Art. no. 9.

[7] A. Bhandari, M. Feigin, S. Izadi, C. Rhemann, M. Schmidt, and R. Raskar, "Resolving multipath interference in Kinect: An inverse problem approach," in *Proc. IEEE SENSORS*, Nov. 2014, pp. 614–617.

[8] D. Kijima et al., "Time-of-flight in fog using multiple time-gated exposures," *Opt. Exp.*, vol. 29, no. 5, pp. 6453–6467, Mar. 2021.

[9] S. Cova, A. Longoni, and A. Andreoni, "Towards picosecond resolution with single-photon avalanche diodes," *Rev. Sci. Instrum.*, vol. 52, no. 3, pp. 408–412, Mar. 1981.

[10] C. Niclass, M. Soga, H. Matsubara, M. Ogawa, and M. Kagami, "A 0.18μm CMOS SoC for a 100m-range 10fps 200×96-pixel time-of-flight depth sensor," in *IEEE Int. Solid-State Circuits Conf. (ISSCC) Dig. Tech. Papers*, Feb. 2013, pp. 488–489.

[11] M.-J. Lee and E. Charbon, "Progress in single-photon avalanche diode image sensors in standard CMOS: From two-dimensional monolithic to three-dimensional-stacked technology," *Jpn. J. Appl. Phys.*, vol. 57, no. 10, p. 1002, Oct. 2018.

[12] C. Bruschini, H. Homulle, I. M. Antolovic, S. Burri, and E. Charbon, "Single-photon avalanche diode imagers in biophotonics: Review and outlook," *Light, Sci. Appl.*, vol. 8, no. 1, p. 87, Sep. 2019.

[13] R. K. Henderson et al., "5.7 A 256×256 40nm/90nm CMOS 3D-stacked 120dB dynamic-range reconfigurable time-resolved SPAD imager," in *IEEE Int. Solid-State Circuits Conf. (ISSCC) Dig. Tech. Papers*, Feb. 2019, pp. 106–108.

[14] O. Kumagai et al., "7.3 A 189×600 back-illuminated stacked SPAD direct time-of-flight depth sensor for automotive LiDAR systems," in *IEEE Int. Solid-State Circuits Conf. (ISSCC) Dig. Tech. Papers*, Feb. 2021, pp. 110–112.

[15] K. Morimoto et al., "3.2 megapixel 3D-stacked charge focusing SPAD for low-light imaging and depth sensing," in *IEDM Tech. Dig.*, Dec. 2021, p. 20.

[16] B. Kim, S. Park, J.-H. Chun, J. Choi, and S.-J. Kim, "A 48×40 13.5mm depth resolution flash LiDAR sensor with in-pixel zoom histogramming time-to-digital converter," in *IEEE Int. Solid-State Circuits Conf. (ISSCC) Dig. Tech. Papers*, Feb. 2021, pp. 108–109.

[17] B. Park et al., "A 64 × 64 SPAD-based indirect time-of-flight image sensor with 2-tap analog pulse counters," *IEEE J. Solid-State Circuits*, vol. 56, no. 10, pp. 2956–2967, Oct. 2021.

[18] B. Park, I. Park, W. Choi, Y. Na, and Y. Chae, "A 40-m range 90-frames/s CMOS time-of-flight sensor using SPAD and in-pixel time-gated pulse counter," *IEEE Solid-State Circuits Lett.*, vol. 3, pp. 422–425, 2020.

[19] D. Cho, B. Park, H.-S. Choi, M.-J. Lee, and Y. Chae, "A 30fps 64×64 flash LiDAR sensor with push-pull analog counter achieving 0.1% depth uncertainty at 70m detection range," in *Proc. IEEE Symp. VLSI Technol. Circuits*, Jun. 2024, pp. 1–2, doi: 10.1109/VLSITECHNOLOGYANDCIR46783.2024.10631325.

[20] R. Schwarte et al., "New electrooptical mixing and correlating sensor: Facilities and applications of the photonic mixer device (PMD)," *Proc. SPIE*, vol. 3100, pp. 245–253, Sep. 1997.

[21] C. S. Bamji et al., "IMpixel 65 nm BSI 320MHz demodulated TOF image sensor with 3μm global shutter pixels and analog binning," in *IEEE Int. Solid-State Circuits Conf. (ISSCC) Dig. Tech. Papers*, Feb. 2018, pp. 94–96.

[22] M.-S. Keel et al., "A VGA indirect time-of-flight CMOS image sensor with 4-tap 7-μ m global-shutter pixel and fixed-pattern phase noise self-compensation," *IEEE J. Solid-State Circuits*, vol. 55, no. 4, pp. 889–897, Apr. 2020.

[23] D. Stoppa, N. Massari, L. Pancheri, M. Malfatti, M. Perenzoni, and L. Gonzo, "An 80×60 range image sensor based on 10 μm 50MHz lock-in pixels in 0.18 μm CMOS," in *IEEE Int. Solid-State Circuits Conf. (ISSCC) Dig. Tech. Papers*, Feb. 2010, pp. 406–407.

[24] D. Kim et al., "Indirect time-of-flight CMOS image sensor with on-chip background light cancelling and pseudo-four-tap/two-tap hybrid imaging for motion artifact suppression," *IEEE J. Solid-State Circuits*, vol. 55, no. 11, pp. 2849–2865, Nov. 2020.

[25] M. Tsutsui, T. Yokoyama, T. Ogawa, and I. Mizuno, "A 3-tap global shutter 5.0μm pixel with background canceling for 165MHz modulated pulsed indirect time-of-flight image sensor," in *Proc. Int. Image Sensor Workshop*, 2021, Paper R17.

[26] K. Hatakeyama et al., "A hybrid indirect ToF image sensor for long-range 3D depth measurement under high ambient light conditions," in *Proc. IEEE Symp. VLSI Technol. Circuits (VLSI Technol. Circuits)*, Jun. 2022, pp. 46–47.

[27] Q. Chen, K. Nie, J. Gao, X. Zhang, and J. Xu, "A 1920 × 1080 array 2-D/3-D image sensor with 3-μm row-time single-slope ADC and 100-MHz demodulated PPD locked-in pixel," *IEEE J. Solid-State Circuits*, vol. JSSC-60, no. 6, pp. 2025–2036, Jun. 2025.

[28] D. Van Nieuwenhove, W. van der Tempel, and M. Kuijk, "Novel standard CMOS detector using majority current for guiding photo-generated electrons towards detecting junctions," in *Proc. Symp. IEEE/LEOS Benelux*, Dec. 2005, pp. 229–232.

[29] Y. Ebiko et al., "Lowpower consumption and high resolution 1280×960 gate assisted photonic demodulator pixel for indirect time of flight," in *IEDM Tech. Dig.*, Apr. 2020, pp. 1–4.

[30] S. Kawahito et al., "CMOS lock-in pixel image sensors with lateral electric field control for time-resolved imaging," in *Proc. Int. Image Sensor Workshop*, 2013, pp. 361–364.

[31] 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 J. Electron Devices Soc.*, vol. 3, no. 3, pp. 267–275, May 2015.

[32] 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.

[33] M.-W. Seo et al., "A 10 ps time-resolution CMOS image sensor with two-tap true-CDS lock-in pixels for fluorescence lifetime imaging," *IEEE J. Solid-State Circuits*, vol. 51, no. 1, pp. 141–154, Jan. 2016.

[34] M.-W. Seo, Y. Shirakawa, Y. Kawata, K. Kagawa, K. Yasutomi, and S. Kawahito, "A time-resolved four-tap lock-in pixel CMOS image sensor for real-time fluorescence lifetime imaging microscopy," *IEEE J. Solid-State Circuits*, vol. 53, no. 8, pp. 2319–2330, Aug. 2018.

[35] Y. Shirakawa, K. Yasutomi, K. Kagawa, S. Aoyama, and S. Kawahito, "An 8-Tap CMOS lock-in pixel image sensor for short-pulse Time-of-Flight measurements," *Sensors*, vol. 20, no. 4, p. 1040, Feb. 2020.

[36] E. R. Fossum, "Active pixel sensors: Are CCDs dinosaurs?" *Proc. SPIE*, vol. 1900, pp. 2–14, Jul. 1993.

[37] E. R. Fossum and D. B. Hondongwa, "A review of the pinned photodiode for CCD and CMOS image sensors," *IEEE J. Electron Devices Soc.*, vol. 2, no. 3, pp. 33–43, May 2014.

[38] J. R. Lakowicz, *Principles of Fluorescence Spectroscopy*. Boston, MA, USA: Springer, 2006, ch. 4.

[39] K. Kitano et al., "Recovering temporal PSF using ToF camera with delayed light emission," *IPSJ Trans. Comput. Vis. Appl.*, vol. 9, no. 1, pp. 2–7, Dec. 2017.

[40] D. L. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.

[41] R. Baraniuk, "Compressive sensing [Lecture Notes]," *IEEE Signal Process. Mag.*, vol. 24, no. 4, pp. 118–121, Jul. 2007.

[42] E. J. Candes and M. B. Wakin, "An introduction to compressive sampling," *IEEE Signal Process. Magazine*, vol. 25, no. 2, pp. 21–30, Mar. 2008.

[43] Y. Hitomi, J. Gu, M. Gupta, T. Mitsunaga, and S. K. Nayar, "Video from a single coded exposure photograph using a learned over-complete dictionary," in *Proc. Int. Conf. Comput. Vis.*, Nov. 2011, pp. 287–294.

[44] T. Sonoda, H. Nagahara, K. Endo, Y. Sugiyama, and R.-I. Taniguchi, "High-speed imaging using CMOS image sensor with quasi pixel-wise exposure," in *Proc. IEEE Int. Conf. Comput. Photography (ICCP)*, May 2016, pp. 1–11.

[45] Y. Luo, J. Jiang, M. Cai, and S. Mirabbasi, "CMOS computational camera with a two-tap coded exposure image sensor for single-shot spatial–temporal compressive sensing," *Opt. Exp.*, vol. 27, no. 22, p. 31475, 2019.

[46] M. Qiao, Z. Meng, J. Ma, and X. Yuan, "Deep learning for video compressive sensing," *APL Photon.*, vol. 5, no. 3, Mar. 2020, Art. no. 030801.

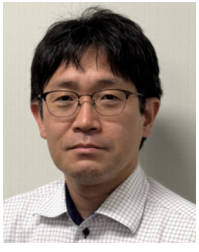
[47] B. Chen et al., "Invertible diffusion models for compressed sensing," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 47, no. 5, pp. 3992–4006, May 2025.

[48] Y. Zhong, C. Zhang, and J. Li, "Image compressed sensing reconstruction via deep image prior with structure-texture decomposition," *IEEE Signal Process. Lett.*, vol. 30, pp. 85–89, 2023.

[49] F. Mochizuki et al., "6.4 single-shot 200Mfps 5×3-aperture compressive CMOS imager," in *IEEE Int. Solid-State Circuits Conf. (ISSCC) Dig. Tech. Papers*, Feb. 2015, pp. 116–117.

[50] F. Mochizuki et al., "Single-event transient imaging with an ultra-high-speed temporally compressive multi-aperture CMOS image sensor," *Opt. Exp.*, vol. 24, no. 4, pp. 4155–4176, 2016.

[51] K. Kagawa et al., "A dual-mode 303-megaframes-per-second charge-domain time-compressive computational CMOS image sensor," *Sensors*, vol. 22, no. 5, p. 1953, Mar. 2022.
[52] A. N. Pham, T. Ibrahim, K. Yasutomi, S. Kawahito, H. Nagahara, and K. Kagawa, "Depth quality improvement with a 607 MHz time-compressive computational pseudo-dToF CMOS image sensor," *Sensors*, vol. 23, no. 23, p. 9332, Nov. 2023.
[53] K. Kagawa et al., "Pseudo-direct LiDAR by deep-learning-assisted high-speed multi-tap charge modulators," in *IEDM Tech. Dig.*, San Francisco, CA, USA, Dec. 2024, pp. 1–4.
[54] M. Yoshida et al., "Single-shot efficient depth imaging based on time-compressive CMOS image sensor," in *Proc. IEEE SENSORS*, Oct. 2024, pp. 1–4.
[55] M. Horio et al., "Resolving multi-path interference in compressive time-of-flight depth imaging with a multi-tap macro-pixel computational CMOS image sensor," *Sensors*, vol. 22, no. 7, p. 2442, Mar. 2022.
[56] T. Okuyama et al., "A short-pulse indirect ToF imager using 6-tap pixel with backside illuminated structure for high-speed charge demodulation," in *Proc. Intl. Image Sensor Workshop*, Jun. 2025, pp. 141–144, Paper R06.5.
[57] R. Miyazawa et al., "A time-of-flight image sensor using 8-tap P-N junction demodulator pixels," *Sensors*, vol. 23, no. 8, p. 3987, Apr. 2023.
[58] C. Li, W. Yin, H. Jiang, and Y. Zhang, "An efficient augmented Lagrangian method with applications to total variation minimization," *Comput. Optim. Appl.*, vol. 56, no. 3, pp. 507–530, Dec. 2013.
[59] S. Boyd, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Found. Trends Mach. Learn.*, vol. 3, no. 1, pp. 1–122, 2010.
[60] J. Ma, X.-Y. Liu, Z. Shou, and X. Yuan, "Deep tensor ADMM-net for snapshot compressive imaging," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 10222–10231.
[61] Z. Meng, S. Jalali, and X. Yuan, "GAP-Net for snapshot compressive imaging," 2020, *arXiv:2012.08364*.
[62] M. Yoshida et al., "Joint optimization for compressive video sensing and reconstruction under hardware constraints," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, Jun. 2018, pp. 1–16.
[63] A. Jarabo, J. Marco, A. Muñoz, R. Buisan, W. Jarosz, and D. Gutierrez, "A framework for transient rendering," *ACM Trans. Graph.*, vol. 33, no. 6, pp. 1–10, Nov. 2014.
[64] B. Bitterli. (2016). *Rendering Resources*. [Online]. Available: https://benedikt-bitterli.me/resources/
[65] R. A. Colyer, C. Lee, and E. Gratton, "A novel fluorescence lifetime imaging system that optimizes photon efficiency," *Microsc. Res. Technique*, vol. 71, no. 3, pp. 201–213, Mar. 2008.
[66] A. L. Paredes, M. H. Conde, and O. Loffeld, "CS-ToF sensing by means of greedy bi-lateral fusion and near-to-optimal low-density codes," in *Proc. 30th Eur. Signal Process. Conf. (EUSIPCO)*, Aug. 2022, pp. 1996–2000.
[67] *The Human Protein Atlas*. Accessed: Oct. 16, 2025. [Online]. Available: https://www.proteinatlas.org



KEIICHIRO KAGAWA (Senior Member, IEEE) received the Ph.D. degree in engineering from Osaka University, Osaka, Japan, in 2001.

In 2001, he joined the Graduate School of Materials Science, Nara Institute of Science and Technology, Ikoma, Japan, as an Assistant Professor. In 2007, he joined the Graduate School of Information Science, Osaka University, as an Associate Professor. Since 2011, he has been an Associate Professor with Shizuoka University, Hamamatsu, Japan. Since 2020, he has been a Professor with Shizuoka University. His research interests include high-performance CMOS image sensors, imaging systems, and biomedical applications.

DAISUKE HAYASHI received the B.E. degree from Shizuoka University, Hamamatsu, Japan, in 2024, where he is currently pursuing the master's degree.

His research interests are computational photography and computer vision.

ARASHI TAKAKURA is currently pursuing the bachelor's degree with Shizuoka University, Hamamatsu, Japan.

His research interests are computational photography and computer vision.

YUTO UMEKI received the B.E. degree from Shizuoka University, Hamamatsu, Japan, in 2024, where he is currently pursuing the master's degree.

His research interests are fluorescence lifetime imaging and computational photography.

MICHITAKA YOSHIDA received the B.E. and M.E. degrees from Kyushu University, Fukuoka, Japan, in 2017 and 2019, and the Ph.D. degree from the Graduate School of Information Science and Technology, Osaka University, Suita, Japan, in 2023.

He has been an Assistant Professor at Okayama University, Okayama, Japan, since 2025. He was a Research Fellow (PD) of the Japan Society for the Promotion of Science, Shizuoka University, Hamamatsu, Japan. His research interests are computational photography and computer vision.

KEITA YASUTOMI (Member, IEEE) received the Ph.D. degree from Shizuoka University, Hamamatsu, Japan, in 2011.

Since 2021, he has been an Associate Professor with the Research Institute of Electronics, Shizuoka University. Since 2024, he has been the CTO of SUiCTE Company Ltd., Hamamatsu, a university spin-off company for CMOS imager developments. His research interests include time-resolved CMOS image sensors and low-noise imagers.

Dr. Yasutomi is a member of ITE and IEICE.

SHOJI KAWAHITO (Fellow, IEEE) received the Ph.D. degree from Tohoku University, Sendai, Japan, in 1988.

Since 1999, he has been a Professor with the Research Institute of Electronics, Shizuoka University, Hamamatsu, Japan. From 2006 to 2023, he was the CTO of Brookman Technology, Inc., Hamamatsu, and since 2024, he has been the Chairperson of SUiCTE Company Ltd., Hamamatsu, a university spin-off company for CMOS imager developments. He has authored and co-authored over 200 articles in peer-reviewed journals. His current research interests include CMOS imaging devices, sensor interface circuits, and mixed analog/digital circuit designs.

YOUNGCHEOL CHAE (Senior Member, IEEE) received the B.S., M.S., and Ph.D. degrees from Yonsei University, Seoul, South Korea, in 2003, 2005, and 2009, respectively.

During his Ph.D. studies, he advanced oversampling ADCs through innovative design techniques, including inverter-based amplifiers. From 2009 to 2011, he was a Post-Doctoral Researcher at Delft University of Technology, Delft, The Netherlands, where he focused on high-precision sensors and interface ICs. He is currently a Full Professor of electrical and electronic engineering at Yonsei University. Since 2012, he has been leading the Yonsei Mixed-Signal IC Group, specializing in innovative analog and mixed-signal circuits and systems for communication, sensing, and biomedical applications. His team's prolific output includes over 150 peer-reviewed articles and more than 50 patents, with over 70 state-of-the-art works published in premier venues, such as IEEE JOURNAL OF SOLID-STATE CIRCUITS, ISSCC, and the VLSI Symposium.

HAJIME NAGAHARA (Member, IEEE) received the Ph.D. degree in system engineering from Osaka University, Osaka, Japan, in 2001.

He was a Research Associate of Japan Society for the Promotion of Science from 2001 to 2003. He was an Assistant Professor at the Graduate School of Engineering Science, Osaka University, from 2003 to 2010. He was an Associate Professor at the Faculty of Information Science and Electrical Engineering, Kyushu University, Fukuoka, Japan, from 2010 to 2017. He was a Professor at the Institute for Datability Science, Osaka University, from 2017 to 2024. He was a Visiting Associate Professor at CREA, University of Picardie Jules Verne, Amiens, France, in 2005. He was a Visiting Researcher at Columbia University, New York, USA, from 2007 to 2008 and from 2016 to 2017. He has been a Professor at the D3 Center, Osaka University, since 2024. His research interests include computational photography and computer vision.