SUMMARY  Time-of-flight (TOF) range imaging is a promising technology for various applications such as touchless control, augmented reality interface, and automotive. The TOF range imagers are classified into two methods: direct TOF with single photo avalanche diodes and indirect TOF with lock-in pixels. The indirect TOF range imagers have advantages in terms of a high spatial resolution and high depth precision because their pixels are simple and can handle many photons at one time. This paper reviews and discusses principal lock-in pixels reported both in the past and present, including circuit-based and charge-modulator-based lock-in pixels. In addition, key technologies that include enhancing sensitivity and background suppression techniques are also discussed.

key words: time-of-flight, CMOS image sensors, lock-in pixel, 3-D imaging, charge modulator

1. Introduction

Over the last two decades, numerous innovations in solid-state imaging devices have changed our lifestyle by making digital cameras more available. This growth has also driven the development of 3D imaging, and many depth cameras are becoming more accessible and pervasive in our lives. Among many 3D imaging technologies, time-of-flight (TOF) range imaging has attracted much attention because of the capability of its small form factor, low cost, and real-time acquisition for various applications such as touchless control, augmented reality interface, and automotive.

TOF range cameras [1] obtain a depth image as well as an intensity image by sensing a round-trip time from a light source to an object, as shown in Fig. 1. The distance, D, is calculated by the time of flight, $t_{TOF}$, and the known light of speed, c, as the following equation.

$$D = \frac{c}{2} t_{TOF}. \quad (1)$$

The TOF ranging system with a scanning mechanism is also called light detection and ranging (LIDAR).

The TOF measurement is divided into two methods: direct TOF (dTOF) and indirect TOF (iTOF) shown in Figs. 2 (a) and 2 (b), respectively. The dTOF uses a short pulse laser and directly measures the TOF using time-measurement circuits such as time-digital converters (TDCs). In order to generate the STOP signal of the time-measurement circuits, single-photon avalanche diodes (SPADs) [2]–[5] are widely used as the detectors of CMOS dTOF imagers. The SPAD is an avalanche photodiode operated in a Geiger mode having an infinity avalanche gain, and generates a pulse associated with an incident photon.
The dTOF imagers exhibit an excellent performance at long distance measurement owing to their high-sensitivity nature [3], [4], [6]. On the other hand, the dTOF imagers require a TDC implementation basically on a unit pixel. The shared TDC architecture [4], [6] relaxes the requirement at the cost of non-simultaneous acquisition for all pixels, and the stacked technology [4], [6] prevents the deterioration of the fill factor of SPADs. However, the spatial resolution is still limited to 65k pixels [4] and 0.1M pixels [6] as dTOF imagers. The recent progress in-depth for SPADs and dTOF imagers is summarized in [7].

Another approach is iTOF measurement that uses a time gating corresponding to a fast electronic shutter, as shown in Fig. 2(b). The iTOF pixels accumulate photo-generated charges within the time gate into in-pixel storage. Such the pixel used in iTOF imaging is called lock-in pixel [8]. The time gate in the lock-in pixels is synchronized with the light, enabling the acquisition of a time-dependent signal (denoted as S in Fig. 2(b)). The change of TOF, \( t_{TOF} \), in time domain is converted to changes in amplitude of the accumulated signal as a function \( f: S = f(t_{TOF}) \), and the TOF is deduced by the inverse function: \( t_{TOF} = f^{-1}(S) \). The readout scheme of the lock-in pixel based on charge modulators, described in Sect. 3, is similar to that of CMOS image sensors. Since the required in-pixel transistors of lock-in pixels are less than that of dTOF imager, iTOF imagers are suitable for higher spatial resolution. Solid-state TOF range cameras with lock-in pixels were launched commercially in the early to mid-2000s by PMD Technology, MESA Imaging, and Canesta. In 2014, Microsoft released Kinect V2 [9] having 0.2M pixels in several hundred dollers for gaming applications. This led to the production of QVGA- and VGA-sized depth cameras by many manufacturers. Recently, Microsoft has released the 1M pixels iTOF camera, Azure Kinect [10]. The launch has been followed by the publication of high-resolution iTOF imagers [11], [12].

The other advantage of iTOF imagers with lock-in pixels is the capability of handling many photons at one time. In outdoor conditions, high background light (BGL) due to the sunlight causes issues of saturation and deterioration of depth precision. The charge-modulator-based lock-in pixel with BGL suppression circuits has the potential to be more resistant to BGL. Furthermore, the feature of handling multiple photon capability also exhibits high depth precision up to 67 \( \mu \)m [13] that is expected to be applicable to 3D scanning systems for component inspection and reverse modeling.

The aim of this paper is to provide an overview of the recent iTOF range imager design. This paper reviews and discusses principal lock-in pixels reported both in the past and present, including circuit-based and charge-modulator-based lock-in pixels. Section 2 describes the basic principle of iTOF range imaging, and the essential element, lock-in pixels, are discussed in Sect. 3. In Sect. 4 describes key technologies for present and next-generation iTOF imagers that enables outdoor use. Finally, the conclusion is presented in Sect. 5.

2. Indirect TOF Imaging

2.1 Modulation Methods in Indirect TOF

The iTOF methods are classified according to the driving method of light source: continuous wave (CW) or short-pulsed (SP) illumination, as shown in Fig. 3. In the CW-iTOF method, a sinusoidal or pulsed illumination with 50% duty is demodulated by time gates with different phases. The phase shift, \( \phi_{TOF} \), is calculated by those accumulated signals. When the light pulse is the sinusoidal with a modulation frequency, \( f_m \), and it is demodulated by four-time phase gates (G1-G4), the phase shift and the TOF are calculated by

\[
\phi_{TOF} = \arctan \left( \frac{N_2 - N_4}{N_1 - N_3} \right)
\]

\[
t_{TOF} = \frac{1}{f_m} \left( \frac{\phi_{TOF}}{2\pi} + k \right)
\]

where \( N_1-N_4 \) are the signal electrons acquired by G1-G4, and \( k \) is an integer. Although the acquired signals contains unwanted signals due to a BGL and unmodulated photogenerated signal, taking difference between two phase signals, i.e. \( N_1 - N_3 \) or \( N_2 - N_4 \), cancels those unwanted component except for their shot noises.

In the actual iTOF imagers using CW-iTOF method, 2-tap pixels [9], [14]–[16] are frequently used instead of 4-tap ones [8], [11], [17]. The 2-tap pixel obtains the four different phase signals using two consecutive frames in which the phase of the light source is changed. The choice of 2-tap lock-in pixels simplifies the pixel design at the cost of halve frame rate and less robustness concerning a motion artifact of moving objects. According to [1], (3) is derived from the definition of discrete fourier transform (DFT). In DFT with \( N \) sample points, \( N/2 - 1 \) discrete frequency component is only deduced. In the ideal CW-iTOF, a single frequency is used. This means that \( N = 4 \), i.e., \( N/2 - 1 = 1 \), is sufficient to deduce the phase delay. Therefore, a multi-tap lock-in pixel of more than 4 taps is unnecessary in CW-iTOF.

Since CW-iTOF measures the phase shift, which has a duplicate value in every \( 2\pi \) cycle, causing phase wrapping that is represented as \( k \) in (3). The phase wrapping causes an ambiguity of depth calculation. This issue comes up to the surface as the modulation frequency increases for higher depth precision. For example, 100-MHz modulation frequency yields an unambiguity range only up to 1.5 m. Thus, an object placed at 2.0 m is misled to 0.5 m. Since Kinect V2 [9] uses a 130-MHz modulation frequency at maximum, solving the issue is essential. A phase unwrapping or ambiguity-resolved technique using multiple frequencies is proposed in [18], [19]. In Kinect V2 [9], three modulation frequencies are used for solving ambiguity. An in-depth description including other phase-unwrapping techniques is summarized in [20].

In SP-iTOF method, the short pulse with a small duty
cycle is used unlike the CW-iTOF. Figure 3 shows an timing example using a 3-tap lock-in pixel (G1-G3) with draining function (GD). The pulse duration of light and time gate are set to be the same ($T_p$). The photogenerated charges within GD set to high are discarded. In the condition of $0 < t_{TOF} < T_p$, the $t_{TOF}$ is calculated by

$$t_{TOF} = T_p \frac{N_2 - N_3}{N_1 + N_2 - 2N_3}$$  \hspace{1cm} (4)$$

Since $N_3$ corresponds to the unwanted signal due to BGL, the denominator $(N_1 + N_2 - 2N_3)$ represents the signal associated with all of the received light except the BGL. Hence, the signal charge ratio, $\frac{N_2 - N_3}{N_1 + N_2 - 2N_3}$, takes a value from 0 to 1, corresponding to the normalized delay with respect to the pulse width. As decrease the duty ratio while keeping average light power, i.e., increasing the peak power, the robustness to the ambient light increases because the acquisition of the BGL component becomes small in the SP-iTOF. Similar to the CW-iTOF method, 2-tap lock-in pixels using two consecutive frames perform the TOF calculation while removing the BGL component. 3-tap lock-in pixels enable the TOF calculation, excluding the BGL component in a single frame.

Unlike in the CW-iTOF, the ambiguity issue is avoidable to set the duty cycle to be small. For example, a condition of 10-ns pulse with 5% duty results in 200-ns cycle time, corresponding that ambiguity occurs every 30 m that does not matter in reality. The short-pulse width leads to better depth precision but decreases the measurable range. To extend the measurable range, a range-shift technique using multiple frames is presented [21]. In the technique, the trigger delay of the light source is varied in several frames, which is equivalent to shifting of the measurable range. The technique allows the expansion of measurable range while keeping depth precision at the cost of a reduced frame rate. SP-iTOF with multi-tap lock-in pixels [22]–[24] obtains a similar benefit without reducing frame rate, described later.

2.2 Depth Precision

Depth results calculated from iTOF sensor outputs fluctuates due to various noises, and the standard deviation ($1\sigma$) is called as depth precision, uncertainty, or range resolution.

The theoretical depth precision can be derived using the error propagation equation to (3), (4). Many works of literature describe theoretical precisions under their condition, e.g., [1], [25]. Here, we would like to give a basic understanding of the dependency of sensor/system parameters on depth precision.

The depth precision, $\sigma_D$, both for CW- and SP-iTOF are given by

$$\sigma_D \propto \frac{D_{MAX}}{C_D} \frac{1}{SNR}$$

$$\propto \frac{D_{MAX}}{C_D} \cdot \frac{\sqrt{N_S + N_{BG} + \alpha N_R^2}}{N_S}$$ \hspace{1cm} (5)$$

where $N_S$, $N_{BG}$ and $N_R$ are a sum of effective signal charges, BGL signal charges, and a dark noise in electrons, respectively. $\alpha$ is a prefactor of dark noise that is determined by readout architecture. The $C_D$ is a demodulation contrast or modulation contrast that is the ratio of modulated signal and offset. The last part in (5) corresponds to $SNR^{-1}$. In its numerator, the first and second terms in square root represent the shot noises due to signal light and BGL, respectively.

The measurable range, $D_{MAX}$, is determined by

$$D_{MAX} = \begin{cases} \frac{cT_o}{2} & \text{if } \text{CW} - \text{iTOF} \\ \frac{cT_o}{2} & \text{if } \text{SP} - \text{iTOF} \end{cases}$$ \hspace{1cm} (6)$$

The dominant factor in precision is different for applications. Here, we consider the following three cases:

- **Case 1**: high light condition with less BGL: non-mobile, indoor use, and short-range application. This case gives an assumption: $N_S >> N_{BG}, N_R$, resulting in $SNR = 1/\sqrt{N_S}$. The depth precision at this condition is intrinsic limit of TOF range measurement. For better precision, reducing $D_{MAX}$ or increasing $N_S$ is required. Since the improvement of precision by increasing $N_S$ is proportional to the square root of $N_S$, reducing $D_{MAX}$ is efficient. Reducing $D_{MAX}$ corresponds to high modulation frequency in CW-iTOF or short pulse width in SP-iTOF. While the $D_{MAX}$ can be extended; the phase wrapping issue due to high modulation frequency is solved using the multiple frequency techniques, and the limited $D_{MAX}$ is extended by range shift technique using multiple frames or multi-tap lock-in pixels in the SP-iTOF.

- **Case 2**: limited light power condition with less BGL such as mobile applications at indoor use. This case suppose that $N_{BG} \approx 0$ and $N_R^2$ comparable to $N_S$. Thus, the dark noise associated with the readout circuits becomes a dominant factor in the depth precision. Obtaining lower dark noise, kTC noise canceling is effective. For this purpose, the implementation of in-pixel intermediate storages is adapted in [10], [17], [26]. Since the choice often reduces the full well capacity, a higher modulation frequency is essential to reduce $D_{MAX}$. Also, [9], [12] implements a binning function to obtain better precision at the cost of reducing spatial resolution.

- **Case 3**: high BGL condition such as outdoor use.
In this case, the reduction of BGL is necessary for better depth precision. In this condition, SP-iTOF method obtain better performance using a smaller duty cycle. Another key point is the choise of light wavelength. The spectral irradiance of sunlight becomes lower at longer wavelength due to an absorption of air. For example, the power density at 940-nm is lower than that at 860-nm by a factor of 2.3.

2.3 SP-iTOF with Multi-Tap Lock-in Pixels

Multi-tap lock-in pixels that allow multiple time gates in single capture offer a wide measurable range without any negative effects on depth precision and robustness to motion artifact. Figure 4 shows the SP-iTOF methods using 3-tap and 7-tap lock-in pixels [23]. Here, the measurement range corresponding to one pulse width ($T_P$) is called time zone. In Fig. 4 (a), the depth calculation performs using the signal ratio between $G_2$ and $G_3$ only. On the contrary, the depth calculation using 7-tap lock-in pixels enables the depth calculation for six time zones in a single frame, as shown in Fig. 4. The $D_{MAX}$ of 7-tap lock-in pixels is six times smaller than that of 3-tap lock-in pixels. Hence, 7-tap lock-in pixels gives six times better precision compared to 3-tap lock-in pixel, while keeping the measurable range. In addition to this, the BGL signals are also reduced in the method.

An efficient timing operation, which is called a depth-adaptive time-gating-number assignment, was also proposed in [23]. For the wide-range distance measurement, the difference of reflected photons between distances in short and long causes an issue; since the reflected photons of short-range are much larger than that of long-range, saturation is more likely to occur in short distances. In [23], the number of accumulation for each time gate ($G_1$-$G_7$) were adjusted so that the reflected photons (from targets having the same reflectivity) are equal for all measurement range. The effectiveness of those techniques was clearly demonstrated in [23].

3. Lock-in Pixel

3.1 Histrical Overview

So far, numerous lock-in pixels have been presented. The lock-in pixels are classified into two categories: charge-modulator-based and circuit-based lock-in pixels. In the 1990s, charge-modulator-based lock-in pixels that were fabricated by CCD or CCD-CMOS technology were presented [8], [14], [27], [28]. In the early 1990s, CCDs were predominated as solid-state imaging devices. However, CMOS image sensors had been actively developed at the time because of their attractive feature: high functionality, low power consumption, and cost-effectiveness [29]. For these reasons, charge-modulator-based lock-in pixels in standard CMOS technology were also investigated [30], [31]. Through these developments, solid-state TOF range cameras were commercialized by PMD technologies, MESA imaging, and Canesta in the early and mid-2000s.

On the other hand, circuit-based lock-in pixels were also investigated in the academic field from the early 2000s [32]. Circuit-based lock-in pixels have several strengths: (1) implementable on a standard CMOS process, (2) easy implementation of BGL suppression, that is because the polarity of charge summation is easily switchable. However, the circuit-based lock-in pixels suffer from noise accumulation during repeated accumulation, while the charge-modulator-based lock-in pixels enable noise-free accumulation. Also, the complicated pixel circuits lead to pixel size limitation or lower spatial resolution. As a result the charge-modulator-based lock-in pixels are more prevalent than the circuit-based ones these days.

In the next subsection, we briefly review the circuit-based lock-in pixels. The further detailed discussion, including case studies, is described in [33].

3.2 Circuit-Based Lock-in Pixel

Circuit-based lock-in pixels are summarized in Table 1. In the early stage, the buffer and sampling type was presented [32]. The concept is very simple; after RT (reset) turns off, the photocurrent due to the incident light integrates on $C_{PD}$, and the resultant voltage is buffered and sampled.
Table 1  Summary of circuit-based lock-in pixels. In the figures, RT and SL stand for reset and select signals, respectively. TG including TGADD, TG SUB, TG1 and TG2 stands for time gate signal, corresponding to electronic shutter shown in Fig. 3.

<table>
<thead>
<tr>
<th>Circuit topology</th>
<th>Simplified example</th>
<th>Remarks</th>
</tr>
</thead>
<tbody>
<tr>
<td>Buffer type</td>
<td><img src="image" alt="Buffer diagram" /></td>
<td>• Simple &amp; less circuit component (opamp is not required.)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Low sensitivity (The conversion gain is determined by C_P0)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• No repeated accumulation at pixel level</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• No CDS (kT/C noise remains)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Power hungry</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Presented in [32, 34] where background suppression and repeated accumulation is implemented in column readout circuits.</td>
</tr>
</tbody>
</table>

| SC-integration                    | ![SC-integration diagram](image) | • Sensitivity improved by charge amplification (The conversion gain is determined by C_INT) |
|                                   |                                   | • Repeated accumulation                                                |
|                                   |                                   | • Background suppression can be implemented by fully-differential topology with a dummy photodiode. |
|                                   |                                   | • No CDS (kTC noise remains)                                            |
|                                   |                                   | Fully-differential pixels using a dummy photodiode are presented in [35, 36]. |

| Two-stage SC-integration          | ![Two-stage SC-integration diagram](image) | • Sensitivity improved by charge amplification [38] (The conversion gain is determined by C_INT) |
|                                   |                                               | • Repeated accumulation                                                |
|                                   |                                               | • Background suppression is implemented by fully-differential approach.  |
|                                   |                                               | • True CDS (kTC noise on C_P0 is cancelled.)                            |
|                                   |                                               | • Large power consumption                                               |
|                                   |                                               | • Complicated topology                                                  |
|                                   |                                               | Presented in [37, 38]                                                  |

| Photocurrent modulation           | ![Photocurrent modulation diagram](image)    | • Repeated accumulation is available                                   |
|                                   |                                               | • Low power consumption                                                |
|                                   |                                               | • BGL suppression                                                      |
|                                   |                                               | • No CDS and additional noise due to the bias current                  |
|                                   |                                               | Presented in [39–41]                                                  |

into C_S using TG (time gate) clock. The time gating is determined by the falling edges of RT and TG. In the actual implementation [32], [34], the C_S is followed by the second buffer and another sampling capacitor in order to enable sample and hold operation in the pixel. The pixel is simple and less circuit components compared to other circuit-based lock-in pixels. However, the pixel suffers from low sensitivity, kT/C noises induced at each accumulation, and no repeated accumulation in the pixel level.

In SC-integration lock-in pixels, a switched-capacitor...
integrator is used for in-pixel repeated accumulation. The time window is determined by TG, while the photo-charges are drained by RT on. The actual implementation uses a fully differential with a dummy photodiode presented in [35], [36]. Using the dummy photodiode, one side of the fully differential terminals integrates the photocurrent only due to BGL; the BGL suppression is also available. Furthermore, owing to the charge amplification, the charge-voltage conversion gain is determined by $C_{\text{INT}}$ that is designed to be smaller than $C_{\text{PPD}}$, the sensitivity is improved compared to the buffer-type lock-in pixels. However, the repeated accumulation integrates the kTC noise on $C_{\text{INT}}$ as well as photo-generated signals.

For lower noise, two-stage SC-integration lock-in pixels are useful [37], [38]. In the pixel, the amplified or buffered signal at the first stage is added with switchable polarity into the second stage using $T_{\text{ADD}}$ and $T_{\text{SUB}}$. The second stage is composed of the switched capacitor circuits. In [37], a fully differential buffer is used as the first stage in order to switch the polarity of summation at the second stage. The difficulty of the lock-in pixels is the requirement of many transistors to implement the circuits and power consumption to implement the two amplifiers. In particular, the second stage requires offset calibration to avoid unwanted signals during the repeated accumulation. The calibration circuit is also implemented in the pixel, leading to less fill factor or large pixel size.

The other lock-in pixels based on photocurrent modulation were also presented [39]–[41]. $T_{\text{G1}}$ and $T_{\text{G2}}$ pulse switches the flow of photocurrent ($I_{\text{ph}}$), and those currents are integrated on capacitors: $C_1$ and $C_2$. The concept was presented in [42], [43] in 1999. However, it is difficult to apply it to TOF range imaging because the photocurrent is too small and the loss of the current integration occurs due to the parasitic capacitance at the photodiode ($C_{\text{PD}}$) during the repeated accumulation. For reducing the loss, the bias and regulation circuits shown in Table 1 were presented [41]. The regulation circuit makes $V_{\text{BG}}$ flowing a constant bias current being the same as $I_0$. Because of the constant bias current, the voltage of the photodiode is regulated to a fixed voltage, the loss due to $C_{\text{PD}}$ is reduced. Although the bias current induces a common-mode signal accumulated on $C_1$ and $C_2$, the in-pixel common-mode rejection circuit cancels the common-mode signal, including the current due to BGL. Since the bias current is designed to be small (2 $\mu$A), the pixel has low power consumption. The pixel, however, suffers from kTC noises and noises due to the constant bias current, which accumulates on the capacitors during the repeated accumulations.

3.3 Charge-Modulator-Based Lock-in Pixel

The charge modulator is the device enabling an electronic shutter in a short time. In other words, photogenerated electrons moving into a specified storage region within a short time window in the order of several ns to several tens of ns. They perform noise-free repeated accumulation, unlike the circuit-based lock-in pixels. The charge modulator is also called a (photonic) demodulator or a photonic mixing device (PMD).

Figure 5 summarizes the device structures and their potential diagram of charge modulators. The quantitative comparison for those device structures are interesting but difficult. One of the performance parameter is modulation contrast or demodulation contrast at the same frequency or pulse width. However, the contrast depends not only on charge modulation speed, but also on the pulse shape and wavelength of light source, the shape of gate driving pulse that strongly depends on pixel counts. Hence, this section gives each features of charge modulators.

Figures 5 (a) and 5 (b) show the charge modulators using photogate [8], [30], [44], [45] and photogate with buried channel [14], [15], [25], [46], respectively. Those structures are basically identical to the surface- and buried-channel CCDs. The channel potential varies due to the applied gate voltage, and different voltages are applied to the gates ($G_1$ and $G_2$) to create the lateral electric field from right to left and vice versa. To realize them in CMOS technology, however, applied voltages to the gates should be reduced to around 3 V while over 10 V is used in CCD technology.

In the buried-channel photogate structure, the fringing field at the gate edges is higher than that of the surface-channel because the equivalent oxide thickness becomes thick; thereby, a smooth lateral electric field is easily obtained. Kawahito et al. presented the photogate structure using CMOS technology [25]. The photogate is formed on the field oxide, and an n-type layer is doped under the photogate to build the buried channel. The thick oxide layer and buried channel enhance the lateral electric field, and 15-µm pixel pitch, 336×252-array TOF imager was demonstrated in 2007.

Figure 5 (c) shows another charge modulator using the pinned photodiode with transfer gates [47]–[50]. In the late 1990s and early 2000s, the pinned photodiode [51], [52] that is one of the most important technology in CCD image sensors was also introduced into CMOS image sensors [53]. The pinned photodiode gives attractive features: a low dark current, the capability of complete charge transfer, and a high full well capacity while keeping a wide depleted (sensitive) region. Among them, the complete charge transfer enables noise-free charge modulation. In the structure, the enhancement of lateral electric field for charge modulation is one of important challenges for TOF range image sensors. This is because the lateral electric field is created only by the fringing field due to the MOS transfer gates at the side of the photodiode. The pinned photodiode region has inherently a constant potential when its size is large (>5-10 µm depending on impurity concentration of the epi layer). For example, in [48], 1-tap lock-in pixel with 12 × 12 µm² was presented; however, the modulation frequency used is only 5 MHz. Kim et al. present another example [50] that has a 6 µm pitch modulator with 70% modulation contrast at 10 MHz in 2012.

Figure 5 (d) shows another pinned photodiode based
structure for high-speed charge modulation using a lateral electric field modulation (LEFM) gate [13], [22], [54], [55]. In the structure, modulation gates are formed along with the pinned photodiode. The gates vary hole concentration of the surface p-layer of the photodiode, creating a lateral electric field in the channel region. The LEFM structure is conceived from a draining-only modulator (DOM) [56]–[59] that has a modulation gate along with the pinned photodiode. Since both LEFM and DOM structure has no transfer gates on the channel that is the path of signal charges, the smoothly-gradient potential is created, resulting in high-speed charge modulation. There are a wide variety of implementation was presented: a 3-tap with drain [13], [55], 4-tap LEFM with drain [22], 2-tap and 4-tap LEFMs with storage diode [60], [61], and 8-tap (or 7-tap with drain) LEFM [23]. In [22], the 4-tap LEFM with 77% modulation contrast at 20-ns gate pulse is achieved.

Figure 5 (f) shows a tapped photodiode (TPD) modulator [65]. Unlike the standard pinned photodiode, surface p+ layers are isolated from the substrate by n-layers such as n1 and n2. Each of these p+ layers has its own electrode, and by applying high and low voltage, a lateral electric field is created. Unlike the modulators with the transfer and LEFM gates, the structure directly modulates the channel potential at the central region of the photodiode, and high-speed carrier transfer is obtained. In [66], a 4-tap TPD modulator has demonstrated a high modulation contrast over 90% with an 8-ns light pulse (850-nm wavelength).

Static-field photodiode structure [67] is preferred to a large pixel design. One of the example is shown in Fig. 6 [68]. At a photogenerated charge at the photodiode is quickly transferred to the modulation diode (MD) and then is transferred to FDs by the transfer gates (G1-G4). In [68], 4-tap pixel having 11.2 µm modulator pitch demonstrated a 4-ns pulse modulation. The concept was first introduced in [67] that uses multiple photogates in 2006. The photogates is biased by multiple voltages generated by a resistor.
ladder. Using a pinned photodiode with various engineered widths, high-speed carrier transfer structures [47], [69] were presented in 2009 and 2010. Recently, Lee et al. presented trident-shaped pinned photodide [70], which demonstrated a modulation contrast of 61.2% at 100-MHz modulation frequency [70], [71] using BSI technology.

Charge modulators using a vertical overflow drain shutter were presented [72], [73]. The VOD shutter was commonly used in interline-transfer CCD image sensors. The VOD shutter acts as a drain (inverse of time gate) and transfer gates are used to acquire signal charges. Although 4-tap readout is implemented [73], one-tap signal is only obtained in single exposure. However, high-speed shutter operation up to 5 ns is achieved with high modulation contrast. The structure is also effective to reduce gating clock skew because of its relatively small load capacitance [52].

Recently, lock-in pixels having a small size of 2-3 µm pitch [10]–[12], [26] were presented for a high spatial resolution iTOF sensor having 1-1.2M pixels. Those designs aim to cover a wide field of view and target a distance up to 5 m, mainly for indoor applications. Those designs have binning function such as 2x2 in [10], 4x4 in [12]. Such sensors with small pixel size require advanced fine-pitch technology such as 65nm BSI [10], 65nm/45nm BSI stacked CMOS [12], 65nm BSI stacked technology [11].

### 3.4 Requirement of Impulse Response for Charge Modulators

When the impulse response of charge modulators is modeled as a Gaussian function, the modulator response, \( h(t) \), is expressed by

\[
h(t) = \frac{1}{\sqrt{2\pi}\sigma_{cm}} \exp \left\{ -\frac{t^2}{2\sigma_{cm}^2} \right\}
\]

where \( \sigma_{cm} \) is an intrinsic response time of modulators. Although this assumption is not always valid, a similar response has been observed with the pixels that are dedicated to high-speed modulation [13]. A similar assumption is also discussed in the ultra-high-speed imagers [74]. Figure 7 shows calculation results of modulation characteristics when \( T_P \) and \( \sigma_{cm} \) are set to be 5 ns and 1 ns. The outputs, N1-N4, and their summation \( (N_{sum}) \) are calculated from the convolution of \( I_{ph} \) and gate pulses. The SP-iTOF with the 4-tap lock-in pixel is emulated for the timing.

The photocurrent, \( I_{ph} \) is calculated from the convolution of \( h(t) \) and the light pulse. The outputs, N1-N4, and their summation \( (N_{sum}) \) are calculated from the convolution of \( I_{ph} \). The modulation contrast between the 2 taps \( = \max \left\{ \frac{N_2 - N_3}{N_1 + N_2 + N_3 + N_4} \right\} \) is 78%. The calculated modulation contrast is equivalent to a 2-tap CW-iTOF one where the modulation frequency is \( f_m = 1/(2T_P) \), i.e., 100 MHz for this condition. Figure 8 shows the calculation results as function of \( T_P/\sigma_{cm} \). To obtain the modulation contrast of \( >90\% \), the intrinsic response \( (\sigma_{cm}) \) should be 1/10 than \( T_P \).

The response of the charge modulators is determined mainly by two mechanisms: (1) charge collection from deep substrate and (2) charge modulation due to the lateral electric field. Since the current TOF imagers utilize NIR light such as 940-nm and the size of recent iTOF pixels shrink, the charge collection determined by the vertical electric field is more important. The theoretical limitation of the intrinsic response \( (\sigma_{cm}) \) at BSI imagers is discussed in [74]. When the substrate thickness is the same as the absorption length \( (\delta \text{ in [74], e.g., 43.5 } \mu \text{m at 940-nm wavelength}) \) and the substrate is biased for the creation of a large and constant ver-
tical field of $2.5 \times 10^4$ V/cm at which the carrier velocity is $9.19 \times 10^{-2}$ µm/ps, the intrinsic response ($1\sigma$) is $3.06 \delta \approx 133$ ps at 940-nm wavelength. Although the calculation does not consider a finite modulation time in any lateral directions and the finite response of gate/light pulse, this indicates that charge modulators themselves have a potential to achieve 375-MHz modulation frequency, 90% modulation contrast, and internal QE of 63%.

3.5 Storage Design

The iTOF imagers basically require global shutter (GS) operation unlike rolling shutter operation, which is used in typical CMOS imagers. The readouts for charge-modulator-based lock-in pixels are typically the same. The FD is typically used as charge storage. Since this configuration yields uncorrelated double sampling for readout, the KTC noise at FDs remains. In the same manner as two-stage charge-transfer GS pixels [75], [75], [76], implementing in-pixel charge storages allow CDS operation for lower readout noise. Those lock-in pixels were reported in [60], [61], [77] for other lock-in pixel applications such as fluorescence lifetime imaging. As described in Sect. 2.2, the influence of readout noise to depth precision depends on operating conditions. For mainly indoor use, recent TOF imagers having small pixel size [10], [17], [26] implement the in-pixel charge storage, achieving a low noise level of around 3.5 e-. However, those pixels have the limited full well capacity of several ten ke- (e.g., 20ke- in [26]).

For outdoor use in which both high and low BGL conditions exist, both low noise and high dynamic range are important. One of the choices to obtain the performance is double-delta correlated multiple sampling (DD-CMS) [78], [79]. Although it requires a 2x readout sequence in a single frame, a relatively low noise level is obtained. In [79], DD-CMS readout allows 5x noise reduction, and the noise level of 15 e- with a conversion gain of 8.3 µV/e is obtained. Another possibility is a readout using switchable conversion gain such as lateral overflow integration capacitors (LOFICs) [80], [81]. Although the implementation of additional capacitors is required, the stacked CIS technology, especially pixel-parallel connection [12], [76], may resolve the issues.

4. Key Technologies and Discussions

4.1 Enhancing Sensitivity for NIR Region

High sensitivity to NIR light is essential for TOF range imagers because the light source power is limited due to eye safety. Longer wavelength has a smaller absorption coefficient of light; it penetrates deeper Si substrate. When the light is absorbed outside of the depletion region, the photo-generated electrons move around by diffusion mechanism. Although some of them come to the photodiode region, the collection speed is not acceptable in TOF range imagers. Therefore, the TOF sensors should have a thick substrate with wide depleted region both for increased NIR sensitivity and high-speed charge collection.

Figure 9 shows the calculation results of internal QE for three NIR wavelengths: 860, 900, and 940 nm. The depletion length of typical CMOS imagers for color imaging is around 3-5 µm, the resulting QEs of 860 nm and 940 nm are only 14-23% and 6-11%, respectively. For outdoor use, the 940-nm wavelength is preferred to 860 nm because of the sunlight spectrum. For 940-nm wavelength, to achieve a good QE of > 80% requires > 70-µm depletion region. Therefore, expanding depleted regions is a key design point for TOF range imagers.

Using a lightly-doped epitaxial layer (at least in [25], [55]) is common and useful to increase the depletion region, which is available to use only the exchange of wafers. However, the channel potential of the charge modulator is limited to 1-2 V, the expansion of the depletion layer is insufficient to improve the NIR sensitivity.

Using a gradual-doping epitaxial layer [82] is one of the techniques to expand the depletion region. A built-in potential due to the difference of doping concentration en-
larges the depletion region. However, the improvement is expected to be small because the potential difference, $\Delta \phi$, of two p-type layers is calculated as

$$\Delta \phi = \frac{kT}{q} \ln \left( \frac{N_{A1}}{N_{A2}} \right)$$

where $N_{A1}$ and $N_{A2}$ is those respective doping concentrations. For example, the condition of $N_{A1}/N_{A2} = 10$ gives only $\Delta \phi \approx 60 \text{ mV}$. In reality, although the structure [82] demonstrates a high-speed modulation of 160 MHz with 55% modulation contrast even at 930 nm, the QE is only 6% even at 850-nm wavelength.

The authors presented an SOI-based charge modulator with a fully-depleted thick substrate [24] in 2020. The conceptual structure is shown in Fig. 10. The modulation gates are formed using active layers on the buried oxide (BOX) that are usually used for a source/drain of SOI transistors. The paper first demonstrates charge modulation capability at 40-ns gate pulse using a fully-depleted thick substrate of $>200 \mu\text{m}$. To the best of the author’s knowledge, it has achieved the highest QE of 55% at a 940-nm wavelength in TOF range image sensors. Since the limited QE is due to the loss caused by the parasitic sensitivity of FDs, further improvement on the QE is expected.

Recently, the authors also have proposed a LEFM-based charge modulator with substrate biasing [79], [83], [84]. The structure has a fully-depleted epitaxial layer with $13-\mu\text{m}$ thickness in the bulk CMOS that is highly compatible with standard CMOS image sensor technologies. The structure theoretically obtains a relatively high QE of 25% in a calculation while keeping high-speed modulation.

Enhancing NIR sensitivity using advanced technologies such as deep trench isolation (DTI) has also been presented [11], [12], [85]–[87]. They implement a scattering [11], [85], [86] or a diffraction [12], [87] structure at the backside surface to increase the angle of incident light. The DTI is also introduced to reduce optical crosstalk between pixels and to increase the QE with multiple reflections to enlarge photon-sensitive length. According to [86], the QE is improved from 20% to 32% at 940 nm using a pyramid-shaped structure for diffraction. In [11], the QE of 38% at 940 nm is obtained using the scattering structure.

### 4.2 BGL Suppression Technique

As demand for outdoor use increases, BGL suppression is becoming more critical. Figure 11 shows calculation results of the estimated photogenerated electrons due to the BGL (110 klx) where the condition is summarised in Table 2. Saturation levels under the various capacitances at the FD are also shown where the available signal swing is supposed to be 1 V. Under no BGL suppression, the saturation occurs immediately, and there is no room to accumulate signal electrons due to the laser pulse.

The basic idea of BGL suppression is to divide the integration in a frame into several sub-integrations. The simplest way is to do a readout for each sub-integration, which is equivalent to increasing the sensor’s frame rate. However, limited readout noise induces at each readout, resulting in lower SNR as well as increasing power consumption. For this reason, in the actual implementation of BGL suppression, a common mode of signal charges or voltages between taps is canceled while the difference signal between taps is stored in pixel or column.

Bamji et.al [9] presented a pixel-level BGL suppression shown in Fig. 12. The concept is similar to [41] shown in Table 1. During the BGL suppression operation ($\phi_{CMR}$ on),

**Table 2** Calculation condition of the photogenerated electrons due to BGL

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wavelength</td>
<td>940 nm</td>
</tr>
<tr>
<td>Optical bandpass filter</td>
<td>40 nm</td>
</tr>
<tr>
<td>F number</td>
<td>1.4</td>
</tr>
<tr>
<td>QE</td>
<td>50%</td>
</tr>
<tr>
<td>Fill factor</td>
<td>50%</td>
</tr>
<tr>
<td>Pixel size</td>
<td>100 $\mu\text{m}^2$</td>
</tr>
<tr>
<td>Reflectivity of a target</td>
<td>0.8</td>
</tr>
<tr>
<td>$T_P$</td>
<td>10 ns</td>
</tr>
<tr>
<td>Signal swing</td>
<td>1.0 V</td>
</tr>
</tbody>
</table>
the signal charges stored in two capacitors (C1 and C2) are summed up with opposite polarity. Therefore the common component due to BGL is canceled, and the differential signals only remain after this operation. This operation can be performed on all pixels at the same time. The advantage of this method is that kTC noises on MC1 and MC2 are also canceled out by taking the difference between taps. Hence the kTC noises are not integrated for each BGL suppression though the kTC noise on MSH remains.

Another approach in pixel-level BGL suppression was presented [88]. The paper proposed a lock-in pixel with a polarity-switchable photodiode and a charge amplifier. The photodiode is composed of p+-n-well-p-sub structure where the p+ and p-sub layers are electrically separated, unlike the standard pinned photodiode. The diode with the p+-n-well enables the treatment of both photogenerated holes and electrons. Those charges are accumulated into the charge amplifier by switching the polarity. Since the dominant sensitivity of photodiode is determined by the n-well-p-sub diode, the BGL cancelation effect is limited. Though, with the proposed diode and sub-frame readout scheme, the paper demonstrates the BGL suppression capability up to 180 klx.

Column-level BGL suppression shown in Fig. 13 is presented in [89]. In the BGL suppression, the difference between taps is taken by an SC integrator implemented in the column readout, and the differential signal is stored into analog memories in the column. After this operation, FDs in the pixel are reset, and another sub-integration begins. The operation is performed row by row. This method does not require a large capacitor in the pixel; it is suitable for small pixels. Obviously, this method requires analog memories as many as the number of pixel rows in every column. For relaxing the requirement, the paper [89] uses 4×4 pixel binning, and a super-resolution technique using shifted binning pixels is used to recover spatial resolution. As a result, the number of required analog memories is reduced to 1/16 of the pixel array. In [71], the analog memories are implemented in each pixel using MIM capacitors. The fill factor is not worse owing to adapting the BSI structure. In column-level BGL suppression, the kT/C noises that are induced at the sampling on analog memories are summed up at each sub-integration. The influence of the kT/C noises on the depth calculation depends on the BGL level; the kT/C noise becomes dominant at a weak BGL while they are negligible at a strong BGL due to its shot noise.

5. Conclusion

In this paper, iTOF range imagers reported over the past two decades are reviewed and discussed to give an overview of iTOF sensor design. The performance of TOF cameras has improved remarkably in terms of spatial resolution and depth precision. For more applications, outdoor and long-distance imaging and low power consumption characteristics will be important. As for coming decade, the continuous evolution of this sensor technology is highly expected.

Acknowledgments

This study is partly supported by the following: (1) KAKENHI No. 19H02194, 18H05240, (2) the Center of Innovation Program from JST, JPMJCE1311. The authors also thank Dr. Keiichiro Kagawa and Mr. Nobukazu Teranishi for a valuable discussion.

References


Keita Yasutomi 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. He is a member of the ITE, IEICE, and IEEE. His research interests include time-resolved CMOS image sensors and low-noise imagers.
Shoji Kawahito received the Ph.D. degree from Tohoku University, Sendai, Japan, in 1988. He has been a Professor with the Research Institute of Electronics, Shizuoka University, Shizuoka, Japan, since 1999. His current research interests include CMOS imaging devices, sensor interface circuits, and mixed analog/digital circuits designs.