A 0.18-μm CMOS SoC for a 100-m-Range 10-Frame/s 200 $\times$ 96-Pixel Time-of-Flight Depth Sensor

Cristiano Niclass, *Member, IEEE*, Mineki Soga, Hiroyuki Matsubara, Masaru Ogawa, and Manabu Kagami, *Member, IEEE*

***Abstract*—With the emerging need for high-resolution light detection and ranging (LIDAR) technologies in advanced driver assistance systems (ADAS), we introduce a system-on-a-chip (SoC) that performs time-correlated single-photon counting and complete digital signal processing for a time-of-flight (TOF) sensor. At the core of the 0.18-μm CMOS SoC, we utilize linear arrays of 16 TOF and 32 intensity-only macro-pixels based on single-photon avalanche diodes in an original look-ahead concept, thus acquiring active TOF and passive intensity images simultaneously. The SoC also comprises an array of circuits capable of generating precise triggers upon spatiotemporal correlation events, an array of 64 12-b time-to-digital converters, and 768 kb of SRAM memory. The SoC provides the system-level electronics with a serial and low-bit-rate digital interface for: 1) multi-echo distance; 2) distance reliability; 3) intensity; and 4) passive-only intensity, thus mitigating system-level complexity and cost. A proof-of-concept prototype that achieves depth imaging up to 100 m with a resolution of 202 $\times$ 96 pixels at 10 frames/s has been implemented. Quantitative evaluation of the TOF sensor under strong solar background illuminance, i.e., 70 klux, revealed a repeatability error of 14.2 cm throughout the distance range of 100 m, thus leading to a relative precision of 0.14%. Under the same conditions, the relative nonlinearity error was 0.11%. In order to show the suitability of our approach for ADAS-related applications, experimental results in which the depth sensor was operated in typical traffic situations have also been reported.**

***Index Terms*—Avalanche photodiodes, depth sensor, light detection and ranging (LIDAR), rangefinder, range imaging, single-photon avalanche diode (SPAD), single-photon detector, 3-D imaging, time-of-flight (TOF) imaging.**

I. Introduction

ACTIVE safety systems have gained increasing interest among car manufacturers worldwide because several new car assessment programs (NCAPs), which encourage automakers to considerably exceed the minimum safety requirements found in local legislation, have expanded their rating criteria to include these systems [1]. Indeed, the European NCAP, for instance, has been revising its rating criteria to focus on overall vehicle safety—including the safety of pedestrians—rather than the safety of only adult occupants. A particular class of safety systems that are of interest in this context are the so-called advanced driver assistance systems (ADASs). Forward collision warning (FCW), autonomous emergency breaking (AEB), adaptive cruise control (ACC), and pedestrian detection systems, to cite a few, rely on various sensing technologies. Among these sensors, millimeter-wave radar remains the sensor of choice and, in some ADAS designs, radar has been combined with passive optical sensors such as monocular or stereo-vision cameras to achieve the desired functions. More recently, active optical sensors based on the time-of-flight (TOF) principle have been emerging as competitive alternative sensors—or as ancillary sensors—for ADAS applications.

A fast-growing class of light detection and ranging (LIDAR) sensors with imaging capability are the so-called TOF three-dimensional (3-D) cameras [2]–[20]. Arrays of light-emitting diodes (LEDs), or laser diodes (LDs), are utilized in these systems to diffusely illuminate targets in the scene. It seems that substantial interest has been fostered in those sensors due to their simple electrooptical construction and, hence, their potential for low-cost and compact implementations. In those sensors, however, the pixels need to cope with extremely low optical signal intensities due to the strong signal attenuation that occurs in diffused light approaches. While they operate reasonably well over a range of several meters in controlled environments, such as indoors, their performance tends to rapidly deteriorate in more severe ambient light conditions. Another common and markedly important limitation in the sensors described in [2]–[20], especially those based on phase-measurement and/or on the indirect TOF principle, is their inability to cope with multi-echo distance evaluation. Multi-echo is prevalent in most practical traffic scenarios. It typically arises when a particular pixel integrates back-reflected signals from two or more targets at different distances or when the optical signal is transmitted through semitransparent media. While the latter issue is evident when the sensor is imaging through glass objects, it may also originate from nonideal weather conditions such as rain and fog. More importantly, however, multi-echo is typically associated with the edges of foreground target objects that are sampled by the finite angular (pixel) resolution of the sensor. In pixels based on the lock-in [2] or on the indirect TOF principles, e.g., [7], multi-echo signals are convolved into a single and often erroneous distance output.

In our laboratory, we have been investigating superior alternatives to diffused-light sensors in order to enable imaging

Manuscript received April 21, 2013; revised July 08, 2013; accepted August 21, 2013. Date of publication October 18, 2013; date of current version December 20, 2013. This paper was approved by Guest Editor Michiel Pertijs.

The authors are with the Information and Electronics Research Division, Toyota Central R&D Labs, Yokomichi 480-1192, Nakagute, Japan.

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/JSSC.2013.2284352

light detection and ranging (LIDAR) to reliably operate in fully uncontrolled environments under the constraints of cost-effective implementations. In [21], a laser-scanning CMOS TOF depth sensor was introduced. The sensor has enabled real-time and long-range depth imaging based on a single LD and has achieved ranging performance comparable to that of a high-end imaging LIDAR system that utilizes an array of 64 discrete LDs [22]. Superior performance with respect to the diffused-light TOF 3-D cameras has been also achieved by an imaging TOF sensor that takes advantage of a micro-electromechanical system (MEMS) scanner approach [23], [24]. Despite their advantages, laser-scanning sensors may also exhibit their own limitations. Miniaturization, high frame rates, and mechanical robustness are substantially more challenging to achieve in these approaches. Avalanche photodiodes biased above their characteristic breakdown voltage, and therefore operating in the so-called Geiger mode, have been at the core of our CMOS sensors. A key feature of these devices, commonly referred to as single-photon avalanche diodes (SPADs), is their ability to detect low-power optical signals down to the photon level with timing resolutions of the order of 100 ps [25]. Circuit-level rejection of uncorrelated photons arising from outdoor ambient light has been achieved by triggering TOF evaluation upon the spatiotemporal correlation of photons [21], followed by advanced time-correlated single-photon counting (TCSPC) processing. So far, however, real-time TCSPC-based digital signal processing (DSP) algorithms, which involve memory-intensive histogram processing, have only been implemented on external field-programmable gate array (FPGA) devices. The separation between the SPAD-based front-end and the DSP back-end integrated circuits, the former typically comprising an array of high-throughput time-to-digital converters (TDCs), has important implications on the system architecture and associated costs. Indeed, in addition to the need for a companion FPGA in the system, the readout interface between front-end and back-end integrated circuits has been characterized by data rates on the order of several gigabit/s [10], [21], [23], [24], thus further increasing system complexity and cost.

In this paper, we report on the design and characterization of a CMOS system-on-a-chip (SoC) that performs complete TCSPC sensing and DSP for a 100-m-range TOF sensor [26]. Multiple targets are resolved per pixel, thus considerably expanding the 3-D point-cloud data available for ADAS-specific computer vision algorithms, while suppressing related erroneous distance evaluations that typically occur in lock-in and indirect TOF pixels. The on-chip DSP also determines the level of integrity of each one of the multi-echo distance evaluations, and simultaneously outputs this reliability information to the system-level electronics. Further enhancement of the sensing data is achieved on-chip by acquiring active (signal and ambient light) as well as passive (ambient light only) intensity images. The CMOS chip therefore provides the system-level electronics with a serial and low-bit-rate digital interface for: 1) multi-echo distance; 2) multi-echo distance reliability; 3) active intensity; and 4) passive intensity, thus mitigating system-level complexity and cost. A six-facet polygonal mirror rotating at 10 revolutions/s combined with a single LD performs laser scanning with a horizontal field-of-view (FOV) of 55 degrees, while achieving a vertical FOV of 9 degrees. Real-time and accurate range images were obtained outdoors in daylight with a resolution of 202×96 pixels, at 10 frames/s, and within a range of 100 m. The worst-case precision obtained with a noncooperative target exhibiting a controlled reflectivity of 9%, which was utilized to model the lowest reflectivity potentially found on common objects, was 0.14%, throughout the measurement range.

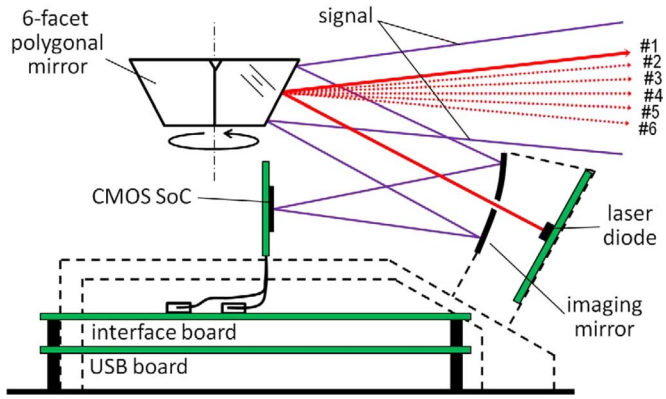

Fig. 1. Overall electrooptical sensor architecture. An 870-nm LD emits pulses of 4 ns FHWM with a repetition rate of 133 k pulses/s. The laser beam is coaxially aimed at a six-facet polygonal mirror through an opening in the center of an imaging concave mirror. Since each facet of the polygonal mirror has a slightly different tilt angle, when rotating at 10 revolutions/s, the polygon performs laser scanning with an FOV of 55×9 degrees.

This paper is organized as follows. In Section II, we describe our LIDAR system design and introduce the new look-ahead passive imaging concept. We then introduce, in Section III, the original DSP algorithm that is implemented on-chip. In Section IV, we cover the sensor chip design in detail. Experimental results are reported and discussed in Section V. The paper is then concluded in Section VI.

II. Imaging LIDAR System

The imaging LIDAR system is based on the laser scanning approach introduced in [21]. It was nevertheless redesigned to implement additional functions and to achieve better performance, and yet be considerably smaller. A diagram of the electrooptical system is shown in Fig. 1. The single LD is aimed coaxially at a six-facet polygonal mirror and emits 4 ns full-width at half-maximum (FWHM) pulses with a repetition rate of 133 k pulses/s. The LD center wavelength is 870 nm, whereas its linewidth is approximately 4 nm FWHM. Its optical beam has a divergence of 1.5 and 0.05 degrees in the vertical and horizontal directions, respectively. Similarly to [21], each facet of the polygonal mirror has a slightly different tilt angle. As a result, in one revolution of 100 ms, the polygonal mirror reflects the laser beam in six vertical directions, thus covering, together with the laser vertical divergence, a contiguous vertical FOV of 9 degrees. During the 55-degree horizontal scanning, at each facet, back-reflected photons from the targets in the scene are collected by the same facet, and are imaged onto the CMOS sensor chip at the focal plane of a concave mirror. An optical near-infrared interference filter, which is not shown in Fig. 1 for clarity, is also placed in front of the sensor to enhance background light rejection. The filter has a passband transmittance and bandwidth of 80% and 10 nm FWHM, respectively, while its stopband transmittance is 1%. Unlike [21], however, the SoC

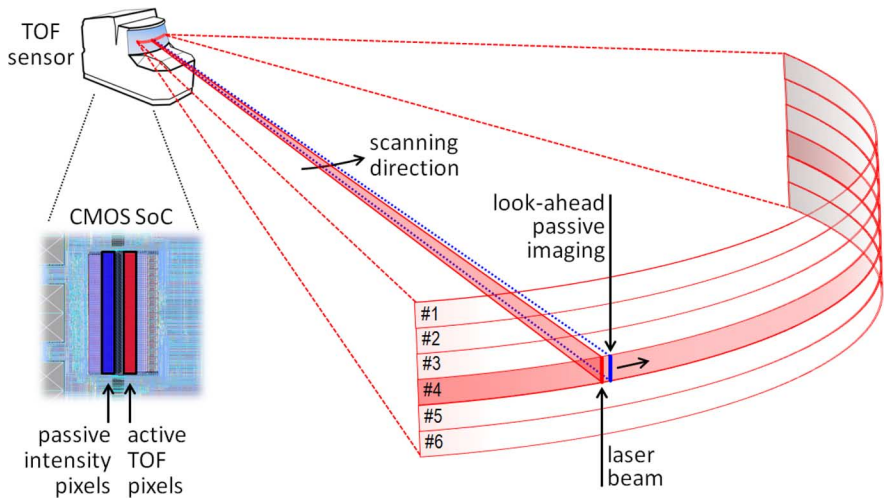

Fig. 2. Diagram of the sensor FOV and proposed look-ahead passive imaging concept. The laser beam is imaged on a linear array of 16 pixels for TOF evaluation, whereas an adjacent linear array of 32 passive pixels acquires intensity information from the scene slightly ahead of the TOF pixels. Taking into account the six-facet scans, the sensor acquires active and passive images with 202 × 96 and 202 × 192 pixels, respectively, at 10 frames/s.

comprises two adjacent vertical linear arrays of pixels. An illustration of the LIDAR system and its FOV is depicted in Fig. 2. One array of 16 pixels is utilized to measure the TOF of the active laser beam, whereas a second array of 32 pixels is utilized to perform passive imaging. The pixels of these arrays resolve different vertical portions of the scene at different facet times, thus generating actual vertical resolutions of 96 and 192 pixels for active images and passive images, respectively. At the end of a complete revolution, active images of 202 × 96 pixels as well as passive images of 202 × 192 pixels are acquired simultaneously. As illustrated in Fig. 2, passive imaging is achieved by an original look-ahead concept, whereby passive pixels are exposed to photons from portions of the scene that have yet to be illuminated by the laser beam. Passive pixels therefore image the ambient-light-illuminated scene at a slightly advanced time with respect to the TOF pixels. Pixel-to-pixel correspondence between TOF and passive images is inherently achieved by design in our concept, thus obviating the need for computationally expensive processing to achieve this function. In addition, since the delay between the passive and active imaging of any portion of the scene is extremely short, 88 μs in this work, typical moving objects in the context of traffic situations do not cause scene mismatches between these images. Passive images can therefore be utilized to accurately predict the amount of background light that is to be integrated on any pixel of the TOF image. The passive images also provide valuable information to the high-level application-specific computer vision algorithms.

The system electronics comprise several printed circuit boards to accommodate different subsystems such as the SoC itself, a laser driver circuit, interface electronics, power supplies, and a universal serial bus (USB) transceiver. The system requires a USB cable for data transfer and a single external ac adapter to operate.

III. Spatiotemporal-Correlated Digital Signal Processing

As described in Section II, a first level of rejection of stray ambient light is achieved in our sensor by the optical near-infrared interference filter. However, background light (BG) suppression at the optical level is insufficient even in moderate daylight conditions due to the overwhelming disparity in power between solar ambient light and the single LD. In order to cope with strong ambient light conditions, the sensor has to perform additional BG rejection at the circuit and DSP levels. In this work, the rejection of BG photons is achieved by an original principle whereby the TOF of signal photons is measured upon the spatiotemporal correlation of two or more photons [21]. The underlying principle is illustrated in Fig. 3. In the figure, the incident optical waveform within a given pixel is depicted as a series of photonic events, which, in turn, are illustrated with wave-packet-shaped symbols. The black (dashed) symbols represent detected photons within a pixel from the uncorrelated ambient light whereas the red (solid) symbols represent signal photons from the LD. In addition, the beginning of each waveform coincides with the emission time of the laser pulses. Since the laser pulse duration is 4 ns, the pixel circuit can purposely ignore individual photonic events that are dispersed in time by at least 4 ns. These isolated photonic events are more likely to originate from uncorrelated background light, which exhibits disparate arrival times. Signal photons are restricted in time to the duration of the laser pulse and, hence, exhibit a considerably higher probability of correlation. The pixel circuit therefore generates trigger signals for TOF evaluation upon the detection of time-correlated photons, i.e., events of two or more photons coincidently detected within 4 ns. As will be described in Section IV, the detection of spatiotemporal-correlated events is achieved by macro-pixels comprising several spatially-confined SPADs.

Under strong ambient light, false correlation events involving two or more uncorrelated photons may still occur at a non-negligible rate. However, these false correlation events are also time-uncorrelated with respect to the laser signal. As a result, in order to further increase the rejection to these unwanted events, a time-correlated DSP is performed at the pixel level. As shown in Fig. 3, each correlation event is saved in a time histogram by adding the number of correlated photons to a histogram bin that is proportional to the measured time delay with respect to the laser pulse emission. The time delay is measured by pixel-level TDCs, as will be described in Section IV. This process is then repeated for several laser pulses. As illustrated

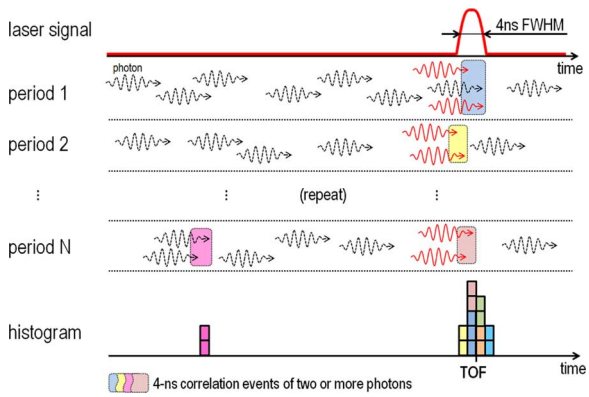

Fig. 3. Diagram of the spatiotemporal-correlated photon counting technique utilized to reject uncorrelated events such as photons from the ambient light. The start times of the depicted waveforms coincide with the laser pulse emission time within each repetition period. The hypothetical detection time of photons within a pixel is illustrated by wave-packet-shaped symbols. The black symbols represent photons from the uncorrelated background light, whereas the red symbols represent photons from the laser signal. Note that despite an incident optical signal-to-background ratio much lower than unity, the resulting histogram suggests that the TOF can be determined reliably.

in Fig. 3, correlated events associated with the laser signal tend to occur with the same delay, thus forming a peak in the histogram around the bins associated with the expected TOF. Since false correlation events exhibit random trigger times, they tend to be uniformly distributed in the histogram. Furthermore, the mean value of the histogram bins associated with false correlation events decreases in proportion to the total number of bins in the histogram.

A typically effective algorithm to extract the TOF of a single-echo return from the histogram data involves determining the location of the bin with the maximum value among all histogram bins [21]. Once acquired, and ignoring second-order effects such as laser speckle and atmospheric turbulence, each bin value of the histogram may be accurately modeled as a Poisson variable. Under extremely low conditions of signal-to-background light ratio (SBR), the stochastic nature of the histogram bins challenges the direct determination of each TOF peak associated with single or multi-echo signal returns. Since the relative fluctuation in a histogram bin with an expected value N decreases with respect to its mean value in proportion to the square root of N, the accurate determination of TOF is improved by accumulating many laser repetition cycles [5]. In the laser scanning approaches, however, such long temporal-averaging process cannot be performed unless image resolution is sacrificed [21]. Another exploration approach to increase the expected values of the histogram bins, and hence, to reduce their relative fluctuations, involves decreasing the number of histogram bins. Indeed, by reducing the number of bins, on average, each histogram bin captures more events, thus increasing its expected value. This technique improves the success rate of discriminating TOF peaks from the uniformly distributed component, provided that the duration of the histogram bin remains comparable to the laser pulse duration. Beyond this critical duration, the contrast between the signal peak amplitude and the mean value of the uniformly distributed component decreases. However, the number of histogram bins is directly related to distance resolution, thus limiting the applicability of this approach.

Algorithm 1

```
RELY(1) := RELY(2) := RELY(3) := 0;
TOF(1) := TOF(2) := TOF(3) := 0:
INT := 0
FOR m ∈ {0, 1, 2, ... , 2047}
   IF h(m) > RELY(1) THEN
      IF m - TOF(1) > 15 THEN
         TOF(3) := TOF(2);
         RELY(3) := RELY(2);
         TOF(2) := TOF(1);
         RELY(2) := RELY(1);
      END IF
      TOF(1) := m;
      RELY(1) := h(m);
   ELSE IF h(m) > RELY(2) AND m - TOF(1) > 15 THEN
      IF m - TOF(2) > 15 THEN
         TOF(3) := TOF(2);
         RELY(3) := RELY(2);
      END IF
      TOF(2) := m;
      RELY(2) := h(m);
   ELSE IF h(m) > RELY(3) AND (m - TOF(1) > 15 AND m - TOF(2) > 15) THEN
      TOF(3) := m;
      RELY(3) := h(m);
   END IF
   INT := INT + h(m);
END
```

Fig. 4. Algorithm utilized to extract multi-echo information from the filtered histogram $h(m)$.

In this work, in order to reduce the fluctuations in bin values due to shot-noise, our proposed DSP algorithm filters the acquired histogram with a finite impulse response (FIR) filter prior to determining the locations of the TOF peaks. This technique significantly improves the success rate of determining actual TOF signal returns without impacting distance resolution. An analysis taking into account the optical specifications as well as the SoC design revealed that histograms with 2048 bins of 11-b words represented a good tradeoff between sensor performance and chip size. The FIR filter consists of a low-pass filter with a 15-word kernel defined as

$$K_{\mathrm{FIR}} = \{6, 8, 10, 11, 12, 13, 14, 15, 14, 13, 12, 11, 10, 8, 6\}. \tag{1}$$

The length and values of K_{FIR} were obtained through simulations of our sensor system model. The filter operation is performed by convolving the histogram with the filter kernel, i.e.,

$$h(m) = \sum_{i=0}^{i<15} h_{\mathrm{RAW}}(m-i) \cdot K_{\mathrm{FIR}}(i) \tag{2}$$

where $h(m)$ is the mth filtered histogram bin with $m \in \{0, 1, 2, \ldots, 2047\}$ and h_{RAW} denotes the unprocessed bin values. The next DSP operation consists of determining the TOF positions of signal peaks in the histogram that correspond to multi-echo signal returns. This is achieved by using the algorithm described in Fig. 4. After a simple initialization, the

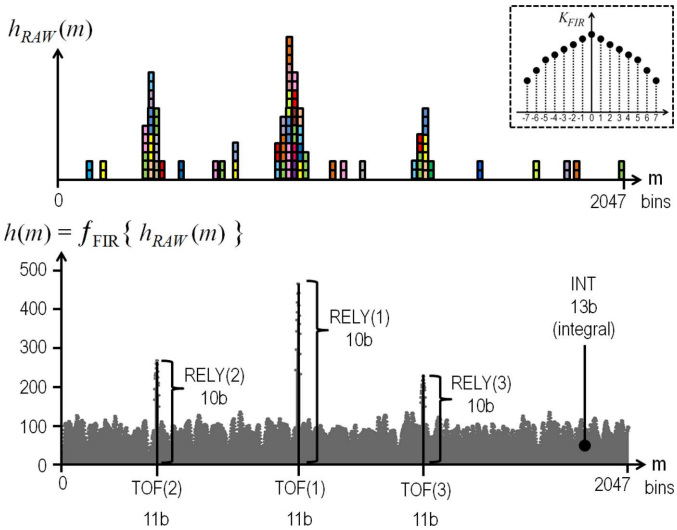

Fig. 5. Idealized representations of the unprocessed and filtered histograms. Inset: plot of the FIR filter kernel. The results of the DSP algorithm are also depicted on the filtered histogram. When combined, these results form a digital word of 76 b per TOF pixel. Note that only the DSP results are actually read out off-chip, thus considerably reducing the system-level readout data rate.

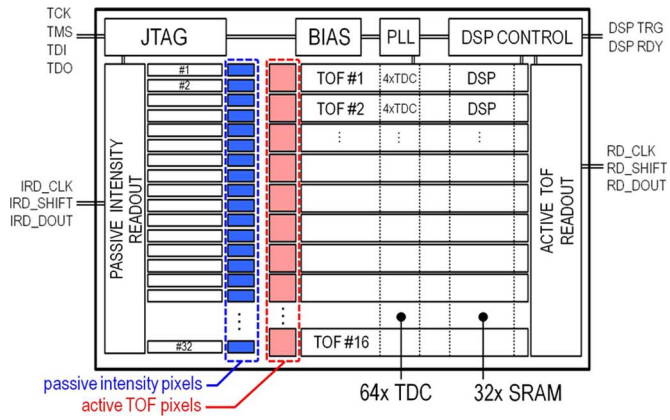

Fig. 6. System-on-a-chip architecture. The chip comprises two linear arrays of pixels: one array is used for passive imaging of the scene, whereas the second one is utilized for TOF imaging. The 16 TOF pixels share an array of 64 TDCs as well as an array of 32 SRAM blocks comprising 768 kb in total. The sensor chip also includes a reference PLL, bias circuitry and an on-chip JTAG controller.

algorithm process the filtered histogram bins and keeps track of the values and positions of the three bins with the three highest values, respectively. However, to prevent these maximum values being related to a single TOF peak, their positions have to be separated by a predetermined number of bins. We assume that the length of the FIR filter kernel, i.e., 15, sufficiently discriminates independent TOF peaks. The algorithm in Fig. 4 utilizes three variables to store the TOF positions of three multi-echo distances, namely TOF(1), TOF(2), and TOF(3), and three additional ones to store their respective amplitudes, i.e., RELY(1), RELY(2), and RELY(3). The latter variables are also utilized as a measure of reliability, from which confidence levels can be determined in the high-level application-specific signal processing. Finally, the DSP algorithm also determines the integral of the overall histogram as a measure of actively-illuminated intensity. This information is saved in a variable called INT, which completes a set of seven processed results that are outputted by the DSP circuit per pixel. Idealized representations of the unprocessed and filtered histograms, as well as of the DSP results, are shown in Fig. 5.

IV. Sensor Chip Design

A. System-on-a-Chip Architecture

A block diagram of the SoC architecture is depicted in Fig. 6. The chip sensor comprises the two linear arrays of pixels that are utilized to perform passive intensity imaging as well as active multi-echo TOF imaging. Each pixel of the passive and active imaging linear arrays consists of several spatially confined SPADs. These pixels are referred to as "macro-pixels." While multi-SPAD macro-pixels are utilized to perform the spatiotemporal correlation function in active pixels, passive-only macro-pixels also take advantage of several SPADs in order to increase their dynamic range. TOF macro-pixels comprise 24 SPADs arranged in a rectangular array of 6×4 elements, whereas the number of SPADs in the passive intensity pixels was reduced to 12, or 6×2, in order to double their vertical resolution. The 16 TOF macro-pixels share an array of 64 TDCs as well as an array of 32 SRAM blocks of 2048×11 b each. The latter, which amount to 768 kb of on-chip memory, are utilized in the DSP circuit. The overall SoC also comprises a reference phase-locked loop (PLL), bias circuitry, and an on-chip joint test-automation group (JTAG) controller. The chip readout interface was highly simplified to mitigate the constraints on the system-level electronics. For test purposes, however, the JTAG controller enables read and write access to most of the on-chip registers, including the on-chip SRAM blocks.

In normal operation mode, in which the JTAG interface need not be implemented at the system level, two readout circuits implement digital serial interfaces for passive and active pixels simultaneously. The passive imaging interface outputs data packets of 544 b, i.e., 32 words of 17-b intensity information that corresponds to the 32 passive pixels. 1216-b active TOF data packets, formed by 3 words of 11-b distance, three words of 10-b reliability, and a word of 13-b signal intensity for each one of the 16 pixels, are read out off-chip via the TOF serial interface. The repetition at which these data packets are read out per image frame corresponds to the number of pixel columns that are sampled by each polygon facet, multiplied by the number of facets, i.e., 202×6. As a result, at 10 frames/s, the off-chip data rates amount to approximately 6.6 and 14.7 Mb/s for passive and active TOF imaging, respectively. These data rates are approximately three orders of magnitude slower than in previous works [10], [21], [23], [24].

B. TOF Macro-Pixel Circuit

Fig. 7 shows the schematics of one active macro-pixel circuit. Single-photon detection is enabled by an array of 24 SPADs that are individually connected to an array of 24 front-end circuits. Each front-end circuit outputs a train of digital rectangular pulses of 4 ns duration upon photon detections in its corresponding SPAD. The outputs of the 24 front-end circuits are combined as an input for a concurrence detection (CD) circuit, which, in turn, asserts its output whenever two or more of its inputs are simultaneously active. Since the output of a front-end

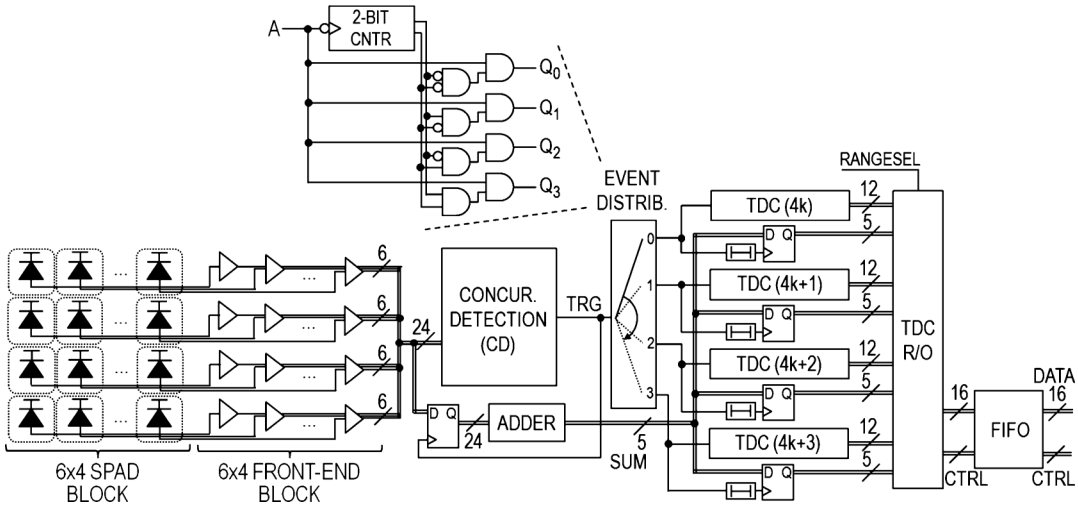

Fig. 7. Schematics of the proposed TOF macro-pixel circuit for the generic kth. Each SPAD in the pixel is connected to its own front-end circuit placed outside the SPAD array. The front-end circuits output digital pulses of 4 ns upon the detection of single photons. These signals are then utilized as the input of a concurrence detection circuit which, in turn, performs the coincidence detection function. TDC conversion time is reduced to 10 ns by distributing trigger events among four TDCs. A local readout circuit combines the TDC results with their associated number of detected photons into 16-b words, which, in turn are written into a sync FIFO.

circuit remains active for 4 ns upon a detected photon, a subsequent detected photon on any other front-end circuit will cause the CD circuit to assert its output, thus achieving the desired function of detecting spatiotemporal correlation events. This macro-pixel circuit architecture, including the design of the CD circuit, was introduced in [21] for a 12-SPAD TOF macro-pixel. In this work, we utilize the same circuit architecture. Nonetheless, the CD circuit was redesigned to detect concurrency of active signals among 24 inputs, instead of only 12 inputs as used in [21].

We also modified the front-end circuit to achieve shorter dead-time and to add a JTAG interface to allow individual SPADs to be disabled for test purposes. The schematics of the front-end circuit are depicted in Fig. 8. With the exception of the JTAG interface, the passive quenching and active recharge circuit operation is the same as that used in [21]. In particular, the D-flip-flop FF1, which maintains the enable/disable state of its corresponding SPAD, is powered by a 1.8-V supply. Its output is conveniently utilized as a cascode bias for a thick-oxide (3.3 V) transistor M_1, which forms a cascode current source with transistor M_2. The current value of this source is designed to be much smaller than the so-called SPAD latching current, thus ensuring proper avalanche quenching [21]. A bias voltage V_{QB} is generated on-chip and is globally distributed to all of the SPADs. The design of the SPADs, including layout as well as the distribution of bias to the common deep n-well cathode (VAPD), is strictly the same as in [21]. As the breakdown voltage (V_{BD}) at room temperature of the SPADs is 22.3 V, an excess bias (V_E) of 3.3 V above V_{BD} is achieved with a VAPD of 25.6 V. Upon photon detection in the SPAD, the avalanche current flows into the cascode current source formed by M_1 and M_2. As shown in Fig. 8(b), since the SPAD current is much larger than the nominal current of this current source, the anode voltage V_{SPAD} rises extremely rapidly, thus reducing the bias across the SPAD towards V_{BD} and eventually quenching its current. Once the SPAD is fully quenched, its junction capacitance as well as any parasitic capacitance on node V_{SPAD} is recharged towards ground by the same current source. The voltage pulse on V_{SPAD}, whose amplitude is approximately V_E, is converted to a rectangular digital pulse by an inverter, shown as V_{INV} in Fig. 8(b). The duration of this pulse approximately defines the SPAD dead time, i.e., the time during which the SPAD is unable to detect subsequent photons, which was set to 20 ns in this work from 40 ns in [21] by adjusting the value of the cascode current source. After a subsequent inversion, the rectangular pulse is shortened to 4 ns by a simple monostable circuit, as shown in Fig. 8(a). The SPAD-level JTAG interface consists of a shadow D-flip-flop (FF2) that mirrors FF1 as well as additional logic to implement usual functions such as capture (CAPT), update (UPD), and shift register (SHIFT).

Referring back to Fig. 7, the remaining functions of the macro-pixel include an adder circuit to sample the number of photons involved in each correlation event, TDCs, and a FIFO memory for synchronization with the subsequent stage, i.e., the DSP circuit. Not shown in Fig. 7, a monostable circuit is added to the output of the CD circuit to generate manageable trigger pulses of 4 ns, thus preventing metastability errors in the subsequent stages. The 24-input of the adder is registered upon CD trigger events, while its output encodes the correlated photon count into a 5-b word. In [21], the maximum conversion rate at which each TDC, and consequently the macro-pixel, could record correlation events was limited to 12.5 Msample/s. System-level simulations of our optical model revealed that the previous conversion rate, which was limited by the off-chip readout throughput, was insufficient under extreme conditions of ambient light outdoors. The requirement for higher conversion rates has also been aggravated by the increase of the macro-pixel size from 12 SPADs in [21] to 24 SPADs in the new pixel. In this work, we leveraged the data compression enabled by the on-chip DSP and increased the maximum event rate of each macro-pixel to 100 Msample/s, thus enhancing the immunity of the TOF pixels to ambient light. This is achieved by distributing correlation events to four pixel-level TDCs. The implemented 12-b TDC array, which was introduced in [27] and expanded to 64 TDCs in [23], takes four cycles of the

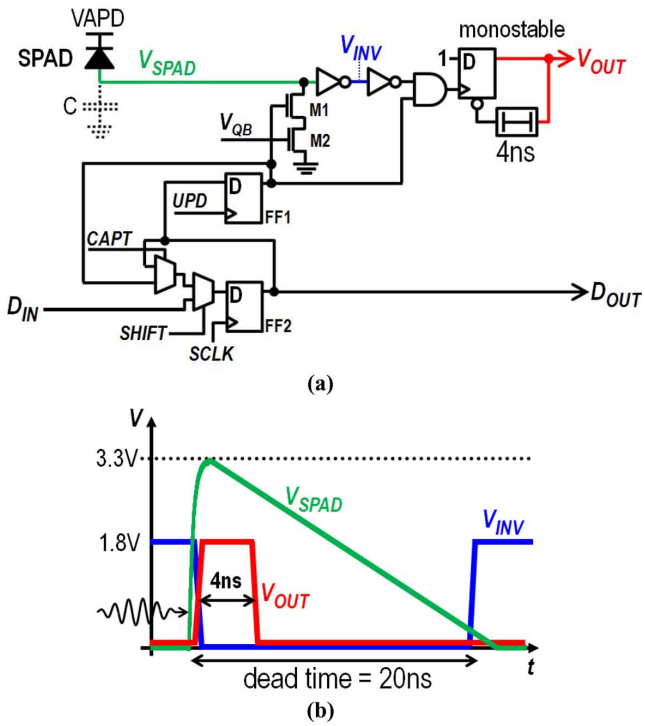

Fig. 8. SPAD and front-end circuit. (a) Passive-quenching and active recharge circuit featuring enable/disable capability via a JTAG interface. (b) Idealized waveforms on nodes V_{SPAD}, V_{INV} and V_{OUT} illustrate the circuit operation when a photon is detected. In normal operation, the output of FF1 is always active. The leading edge of the pulse on node V_{OUT} gives a very precise trigger on the arrival time of the photon, typically on the order of 100 ps.

100-MHz SoC clock to complete a conversion. By sequentially distributing correlation events among four TDCs, the effective conversion time of a macro-pixel is reduced to 10 ns. The event distribution is performed by an AND-based decoder that is driven by a falling-edge-triggered 2-b counter, as shown in the inset of Fig. 7. As can be seen in the figure, the propagation delay from the input A of the decoder to any of its outputs $(\mathrm{Q}_{0,1,2,3})$ is perfectly matched, to the extent that the four output AND gates are matched to one another. A pixel-level TDC readout circuit operating at the SoC clock frequency reads out valid TDC results with their corresponding number of correlated photons, and encodes them into 16-b words that are written into the synchronization FIFO. Since the DSP histogram has 2048 bins (see Section III), it cannot accommodate the full 12-b resolution of the TDCs. The TDC resolution is therefore compressed by one bit by discarding either the most or the least significant bit, via the range selection signal RANGESEL. Nominally, a TDC range of 853.3 ns [27] is utilized, thus increasing its least-significant bit (LSB) resolution to 416.7 ps. This setting corresponds to a resolution of 6.25 cm within a range of 128 m of TOF distance. Optionally, RANGESEL may be set via the JTAG controller to achieve a distance resolution of 3.125 cm within a range of 64 m.

C. Digital Signal-Processing Circuit

The proposed histogram-based DSP algorithm in Section III has the advantage of achieving virtually unique signal discrimination from uncorrelated noise in TOF evaluations. This advantage comes at a cost related to circuit area and processing time. The latter is a critical issue since, similar to other shot-noise limited optical detection systems, the overall performance depends on the ability of the sensor to maximize its exposure to the optical signal. For instance, suspending the photonic signal acquisition of pixels to perform DSP and/or histogram reset would indubitably affect ranging performance or image resolution. In our design, we have prevented the interruption of signal acquisition by integrating two memory blocks of 2048 × 11 b per pixel. These blocks are globally arranged in two memory banks of 16 blocks each, namely, bank A and bank B. Fig. 9 shows a block diagram of the macro-pixel-level DSP circuit. It implements two basic functions independently: histogram acquisition and histogram DSP. The memory banks are assigned successively to histogram acquisition and DSP via a mode-selection signal MSEL. While the histogram acquisition of a given column of pixels on a polygon facet is performed in memory bank A, the DSP of the previously acquired data is performed in memory bank B. Once the pixel acquisition time elapses for that column, the memory banks are swapped and the process is repeated. This approach therefore allows for the simultaneous acquisition and processing of TOF data. As shown in Fig. 9, the FIFO circuit of the macro-pixel is connected to the histogram acquisition circuit, which, in turn, builds a histogram from the spatiotemporal correlated events originating from the FIFO. The histogram DSP circuit, upon completion, registers the multi-echo DSP results as well as the histogram integral onto output registers, which are subsequently serially readout off-chip.

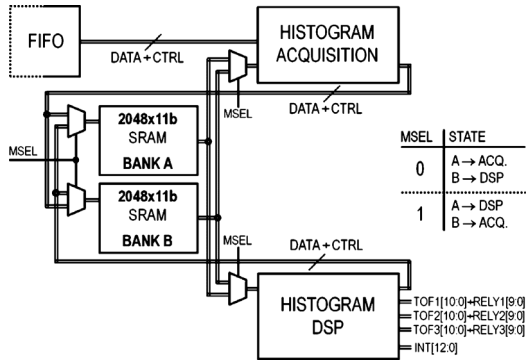

Fig. 9. Block-level diagram of the DSP circuit associated with each TOF macro-pixel. In order to enable the histogram acquisition and the DSP at the same time, every pixel comprises two memory blocks of 2048 × 11 b each. These memory blocks are globally arranged into two memory banks: A and B. First, memory bank A is assigned to histogram acquisition whereas bank B is assigned to DSP circuits. Once the pixel acquisition time elapses, the banks are swapped. The DSP circuits then process the previously acquired histogram data in bank A, while bank B acquires new data.

Fig. 10(a) indicates the coordinates of pixels within the TOF image frame as well as the actual positions of the subsections comprising 202 × 16 pixels that are associated with the polygonal mirror facets. Fig. 10(b) illustrates the waveform of a number of relevant signals related to the image acquisition from the TOF image frame level towards the pixel level. The overall frame acquisition takes 100 ms when the sensor is operating at its nominal frame rate of 10 frame/s. Signal collection cannot be performed when the sensor aperture is aligned with the joining edges of the mirror facets. The effective exposure time of each facet is therefore limited to 8.89 ms, as illustrated in the figure. During the 7.77 ms of inactive time between facets, most of the SoC functions, including pixels and the

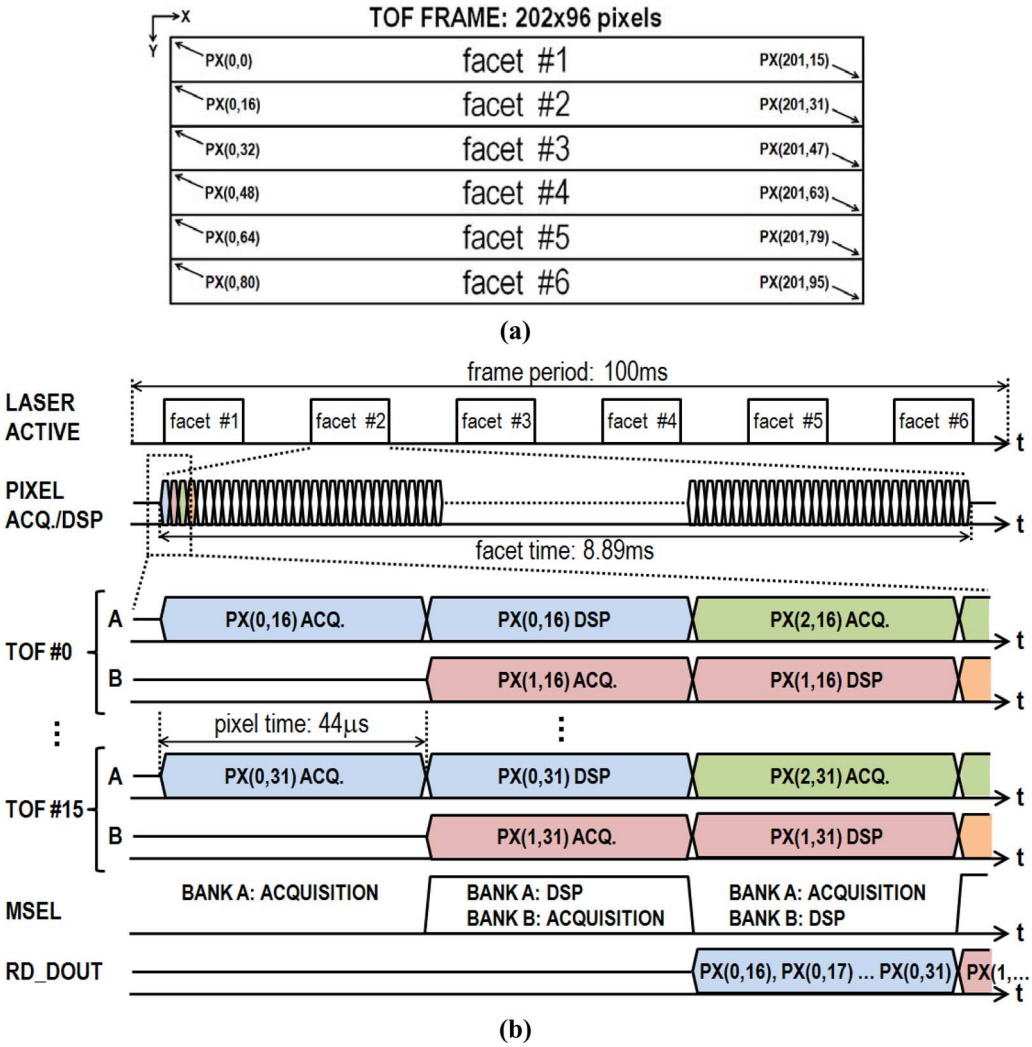

Fig. 10. (a) Pixel coordinates within each polygonal-facet subsection of the TOF image frame. (b) Waveforms of a number of relevant signals related to pixel histogram acquisition and DSP.

DSP circuitry, remain inactive. The LD emission is therefore operated in bursts of 8.89 ms, six times per frame. Each active facet time is subdivided into 202 time intervals during which a sub-column of 16 TOF pixels are acquired and processed in parallel. These intervals, which correspond to the pixel acquisition time, take precisely 44 μs. For instance, facet #2 acquires data for the pixels $\mathrm{PX}\{(\mathrm{x},\mathrm{y}){:}0 \leq \mathrm{x} \leq 201 \wedge 16 \leq \mathrm{y} \leq 31\}$ of the TOF image frame. At the beginning of its active time window, $\mathrm{PX}\{(0,\mathrm{y}){:}16 \leq \mathrm{y} \leq 31\}$ are acquired in memory bank A. Once the pixel acquisition time has elapsed, memory bank A switches into DSP mode and memory bank B starts the acquisition of histograms for $\mathrm{PX}\{(1,\mathrm{y}){:}16 \leq \mathrm{y} \leq 31\}$. In DSP mode, the 16 acquired histograms in memory bank A are digitally processed in order to output distances, reliabilities, and intensities. At the end of the acquisition time of $\mathrm{PX}\{(1,\mathrm{y}){:}16 \leq \mathrm{y} \leq 31\}$, the memory banks are switched again so the acquisition of $\mathrm{PX}\{(2,\mathrm{y}){:}16 \leq \mathrm{y} \leq 31\}$ is performed in bank A. During this time, the serial readout circuit outputs the registered DSP results of $\mathrm{PX}\{(0,\mathrm{y}){:}16 \leq \mathrm{y} \leq 31\}$. This process is then continued until all of the pixels associated with each facet are fully acquired, processed, and read out.

Assuming an empty histogram at the beginning of the pixel time, the histogram acquisition circuit reads the 16-b word data out of the FIFO and utilizes the 11 LSB, i.e., TDC output, as the address signal of either bank A or bank B memory, depending on MSEL. The contents of the memory at that address are read out in one clock cycle. In the following cycle, the memory contents are added to the five most significant bits (MSBs) of the FIFO data, i.e., the photon count, and the result is rewritten to the same memory address. This process is repeated for all of the correlation events within the pixel acquisition time.

The histogram DSP is performed taking advantage of a five-stage pipeline circuit, shown in Fig. 11. The timing diagram associated with the pipeline operation is illustrated in Fig. 12. The DSP execution is started by an initialization operation (INIT) whereby the contents of a number of internal registers are erased: A 15-word register $HB(i)$, where $i \in \{0,1,2,\ldots,14\}$; 3-word registers TOF(n) and RELY(n), where $\mathrm{n} \in \{0,1,2\}$; a single-word register INT, and a memory address register ADDR. As the histogram memory is processed sequentially by incrementing its address, HB is utilized to preserve the 15 trailing bin values of the histogram. TOF(n) and RELY(n) are utilized to register the TOF location and reliability

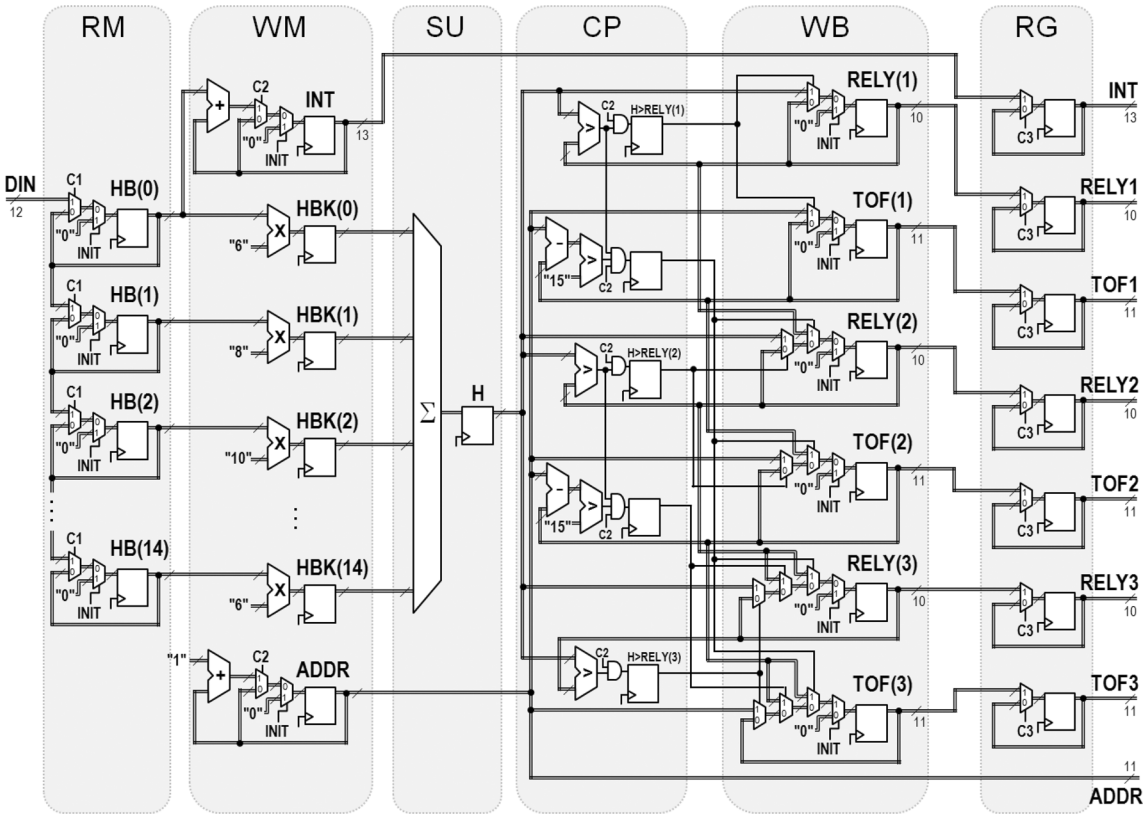

Fig. 11. Simplified register-transfer level schematics of the histogram DSP circuit. The proposed circuit takes advantage of a five-stage pipeline approach. The operations on each histogram bin are labeled as: read memory (RM), write memory (WM), sum (SU), compare (CP), and write back (WB).

of the three TOF echoes, respectively, while INT is utilized as a register to calculate the integral of the histogram. As shown in the diagram in Fig. 12, the pipeline execution starts with a read memory operation (RM). In this cycle, the histogram memory is read at the address pointed to by the ADDR register and the bin value (DIN) is written into the first element of HB, i.e., $HB(0)$. In the same cycle, the elements of HB are simultaneously shifted, i.e.,

$$HB(i) = HB_{\mathrm{prev}}(i-1) \tag{3}$$

for $i \in \{1, 2, 3, \ldots, 14\}$. As a preparation for the next pixel acquisition, the histogram contents have to be sequentially reset. Since our on-chip SRAM blocks have a single-port interface, the reset operation requires another clock cycle at the same memory address. Consequently, in the following clock cycle, denoted as write memory (WM) in the diagram, the memory contents at the same address are erased and ADDR is incremented. In the same cycle, two parallel computations are performed. While $HB(0)$ is added to the existing value of INT, HB is element-wise multiplied with the 15-word filter kernel K_{FIR}, defined in Section III, and the result is stored in a 15-word register (HBK), i.e.,

$$HBK(i) = HB(i) \cdot K_{\mathrm{FIR}}(i) \tag{4}$$

for $i \in \{0, 1, 2, \ldots, 14\}$. In order to complete the convolution computation defined in (2), the next operation involves a sum (SU) operation of all the elements of *HBK*. The result is saved into a register called H as

$$H = \sum_{i=0}^{i<15} HBK(i). \tag{5}$$

In the next cycle, several comparison (CP) operations are performed. The current value of H is compared to RELY(n), where $\mathrm{n} \in \{0, 1, 2\}$, as shown in Fig. 11, while taking into account additional subtleties according to Algorithm 1 listed in Fig. 4. New values of TOF(n) and RELY(n), where $\mathrm{n} \in \{0, 1, 2\}$, are finally written back (WB) to these registers on the final stage of the pipeline, depending on the result of the comparisons performed in the previous cycle. The pipeline execution is then continued for all of the histogram bins, although delayed by two cycles instead of one due to the single-port SRAM access. This feature is rather a design choice than a circuit limitation. Since the histogram comprises 2048 bins, the overall processing, including a final step (RG) to save the DSP results into the output registers, requires 4102 clock cycles. In addition, the DSP circuit operates at the same frequency as the SoC main clock, i.e., 100 MHz. As a result, the overall DSP processing takes 41 μs, which can be effectively accommodated within the pixel integration time of 44 μs. A dual-port SRAM interface could be utilized to reduce the DSP time at the expense, however, of larger circuit area.

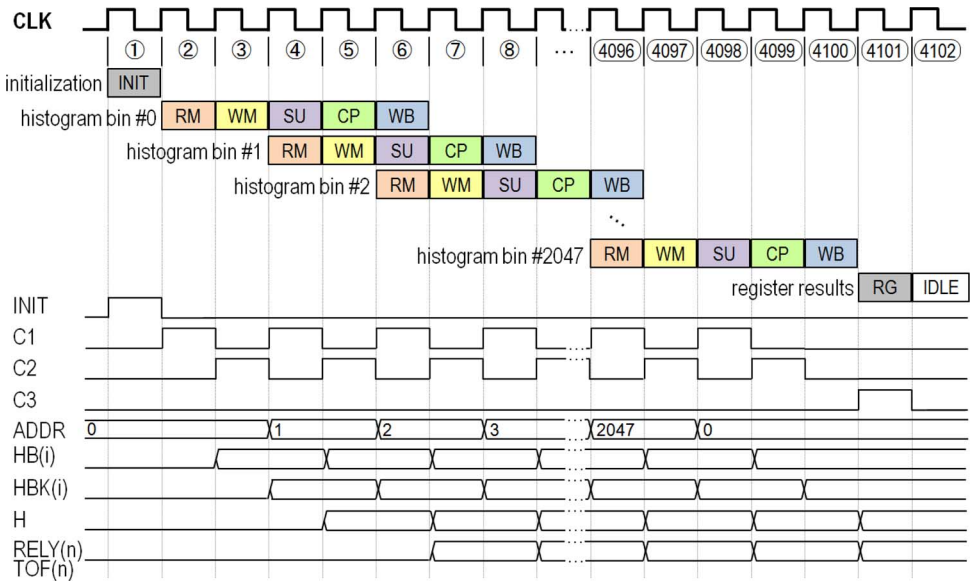

Fig. 12. Timing diagram of the pipeline operations executed in the proposed DSP circuit. Since the actual histogram memory has a single-port interface, the pipeline operations are delayed by two clock cycles.

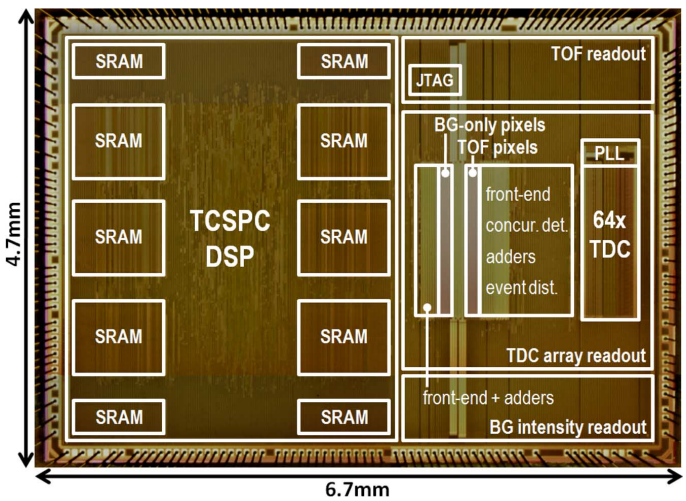

Fig. 13. Photomicrograph of the prototype system-on-a-chip implemented in 0.18-μm high-voltage CMOS.

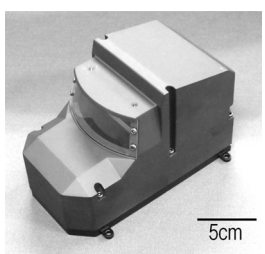

Fig. 14. Assembled proof-of-concept LIDAR system.

V. Experimental Results

A. Sensor Chip and Proof-of-Concept Prototype

A photomicrograph of the prototype chip is depicted in Fig. 13. The photosensitive area, which consists of two linear arrays of macro-pixels, occupies a very small fraction of the 6.7×4.7 mm^2 chip. The chip floor plan was mostly defined by the placement of the proposed DSP circuit, and, in particular, by the 32 blocks of SRAM. The sensor chip was thoroughly tested and confirmed to be fully functional at its nominal operating frequencies. In particular, the operation of the DSP circuit at its nominal operating frequency of 100 MHz has been confirmed at temperatures from 0 °C to 70 °C. The characteristics of the implemented SPADs have been reported in a previous work [21]. The devices exhibit a mean dark count rate (DCR) of 2.65 kHz at room temperature, which, taking into account the SPAD area, corresponds to a mean DCR density of 6 Hz/μm^2. In addition, owing to the physical separation between the SPADs and the front-end circuits, and to multiple-iterations of optimization, the macro-pixels benefit from an excellent fill factor of 70%. The TDC array has also been previously characterized with respect to differential nonlinearity (DNL) and integral nonlinearity (INL) in [23]. The worst-case DNL and INL among all of the 64 TDCs were -0.17 LSB (35 ps) and 0.56 LSB (-116 ps), respectively. The latter figure, if taken alone, corresponds to a distance nonlinearity of 17.4 mm throughout the TDC range of 128 m, thus having a negligible impact on the TOF ranging performance.

A photograph of the assembled proof-of-concept prototype is shown in Fig. 14. The 600-rpm rotating polygonal mirror and the remaining electrooptical components are enclosed in a custom case. When compared with the previous generation [21], the sensor volume was reduced by approximately 50%. In the current state, however, a large margin for miniaturization still exists. The overall system power consumption amounts to 6.4 W when the sensor operates in a typical sensing scenario. The major contributor for the overall power consumption is the LD driver circuit with 3 W, which comprises a 140-V dc–dc step-up regulator, followed by the polygon mirror driver with 1.2 W. In the same conditions, the total SoC power consumption is approximately 530 mW.

B. Qualitative Evaluation of Ranging Performance

In order to investigate the suitability of our approach in ADAS-related applications, we have installed one of the

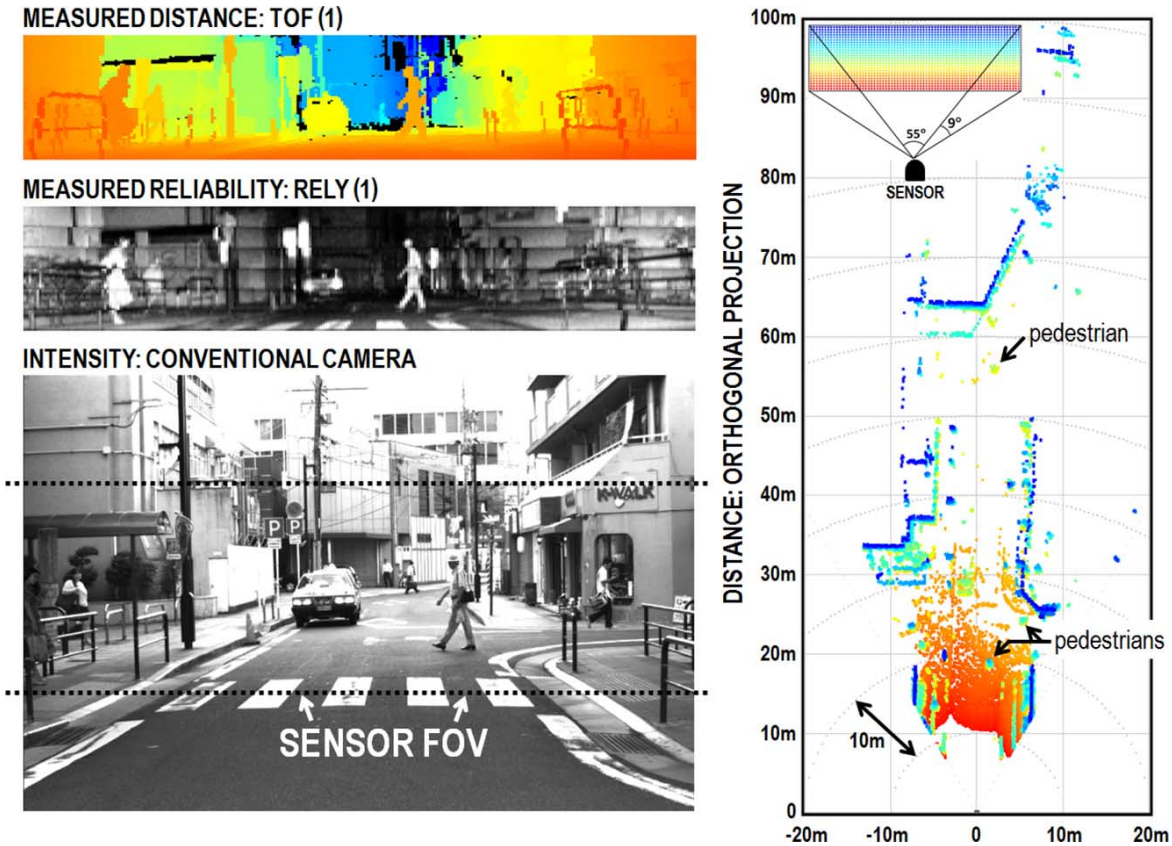

Fig. 15. Distance, reliability, and orthogonal projection of distance point cloud acquired in daylight conditions. The proposed sensor and a conventional camera were mounted on our experimental car. In the orthogonal projection, the vertical position of the pixels is color-coded from red in the bottom row to blue in the top row.

assembled LIDAR sensors on the bumper of an experimental car. On-board computers allow us to visualize and record real-time intensity and multi-echo depth images. In order to obtain reference images of the scene for later analysis, a conventional camera is also mounted on-board near the rear-view mirror and records the scene through the windshield. A snapshot of one image frame acquired with our setup in daylight conditions is shown in Fig. 15. In the figure, the top image on the left-hand side is a color-coded distance image, i.e., TOF(1). In this image, the color is assigned according to the measured depth of each pixel, with near objects appearing in red as opposed to far objects that are depicted in blue. The image in the middle of the images on the left-hand-side corresponds to the measured signal reliability, i.e., RELY(1). The bottom image was taken using the on-board conventional camera. The actual FOV of the LIDAR sensor is illustrated on this image. On the right-hand side of Fig. 15, an orthogonal projection of the measured 3-D point cloud is color-coded on a Cartesian plot. On this plot, the color indicates the row position in the depth image, where red is the bottom row whereas blue indicates the top row. In the orthogonal projection, tall objects therefore appear in blue while the pixels imaging the ground typically appear in red.

Some artifacts are noticeable on the measured reliability image of Fig. 15. Signal nonuniformity is clearly noticeable at the row locations where the facet sub-images join. This nonuniformity is caused by the vertical profile of the laser beam, which, as expected by design, is not perfectly flat. In addition, due to the optical design of the LIDAR system, the vertical orientation of the laser beam tends to vary with the actual polygon angle, which in turn causes distortion towards the horizontal edges of the image. Since these effects do not occur when the laser beam is pointed straight ahead of the sensor, the center portion of the image is inherently free from distortion effects. Distortion effects due to the design of the optical system can be modeled and systematically corrected on each computed 3-D point. This is achieved in real-time in our sensor and, for instance, could be applied to the orthogonal projection image, as shown on the right-hand side of Fig. 15. As can be seen in the figure, not only can pedestrians and cars be readily recognized, but also a number of road features. In particular, narrow obstacles such as distant pedestrians and electric poles, which may be particularly challenging to other alternative sensor techniques, can be sharply imaged by our sensor, even at distances of several tens of meters.

The multi-echo and look-ahead-passive imaging capabilities of the sensor was also verified experimentally in daylight conditions. Fig. 16 depicts acquired images obtained when the sensor was placed in front of a window in our building to image outdoors through the glass. On the left-hand side, the top four images correspond to the color-coded TOF and grayscale reliability images of the first (strongest) echo, followed by the equivalent TOF and reliability images of the second echo. The fifth image from the top corresponds to the passive image obtained with the look-ahead concept. The figure also includes a photograph taken with a conventional camera as a reference image, shown in the bottom of Fig. 16, as well as the orthogonal

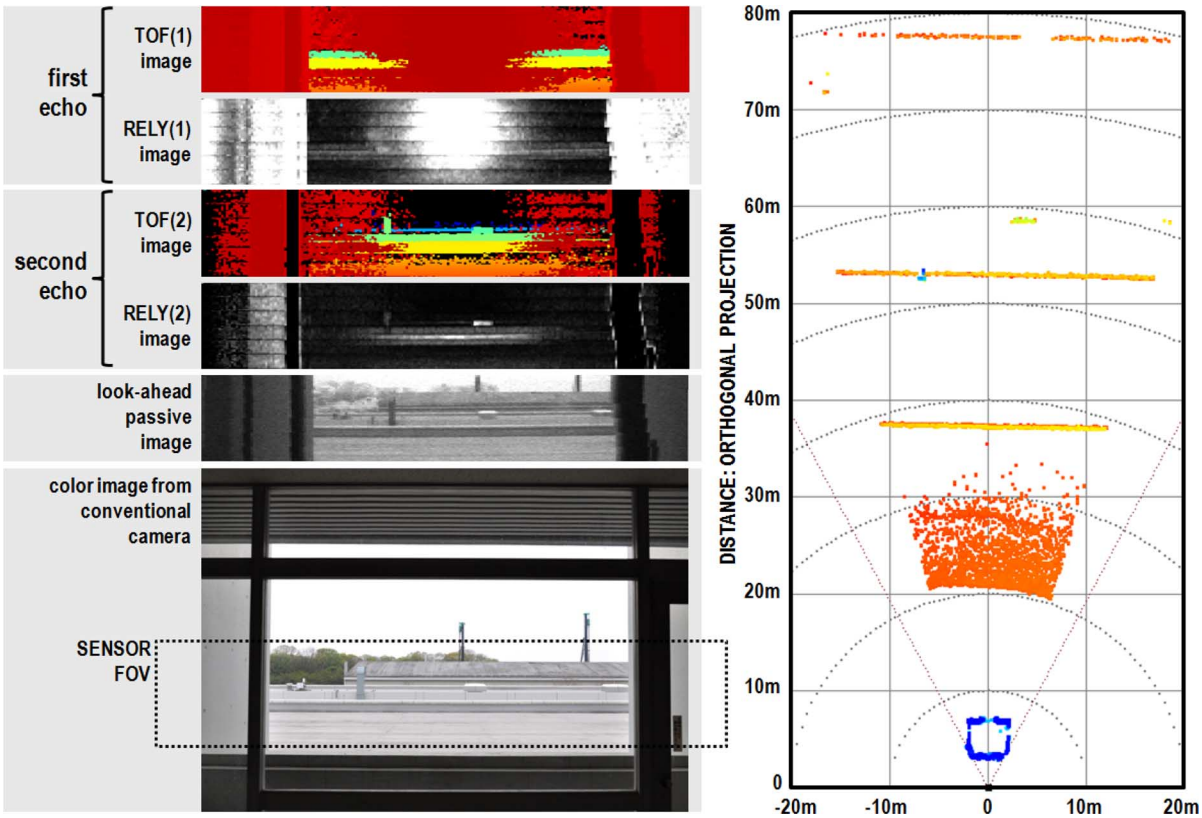

Fig. 16. Multi-echo and passive imaging obtained when the sensor was placed in front of a window in our building to image outdoors. On the left-hand side, the top four images correspond to the color-coded TOF and reliability images of the first (strongest) echo, followed by the TOF and reliability images of the second echo, respectively. The fifth image from top corresponds to the passive image obtained with the look-ahead concept. The figure also includes a reference photograph taken with a conventional camera, shown as the bottom image on the left-hand side. An orthogonal projection plot of the acquired TOF data is shown on the right-hand side. In this plot, the vertical position of the pixels is color-coded from red in the bottom row to blue in the top row.

projection of the acquired TOF data, shown on the right-hand side. Note that, with the exception of the photograph, all the remaining images in Fig. 16 were acquired simultaneously with a single scan when the sensor was operating at 10 frames/s. As can be seen in the reference image, where the sensor FOV is illustrated, the targets in the scene comprise the window and the roof of two adjacent buildings. A specular reflection on the window glass creates a spot that appears saturated in the reliability image associated with the first echo. This specular reflection was very strong and, hence, it back-illuminated the room where the sensor was placed. The back of the room was therefore also imaged by the sensor due to a multi-path situation. Indeed, reflected signals on the wall behind the sensor propagated back to the sensor via an additional reflection on the window glass. As a result, the room layout appears in the distance images as if the room were outside the building, as can be seen in the orthogonal projection image of Fig. 16. This is one of the most challenging multi-path scenarios that TOF sensors have to cope with. The TOF pixels related to the specular reflection on the window glass appear in red in the first echo image, i.e., TOF(1). However, pixels related to the multipath situation appear distributed among the first-echo and second-echo images. Despite the severe multipath situation, the sensor was able to determine the distance of several targets through the window glass and up to approximately 80 m of distance, thus demonstrating its multi-echo capability. Furthermore, the simultaneously acquired passive image also demonstrates the correct operation of the proposed look-ahead concept. When compared to the photograph taken using a conventional camera, the passive image exhibits higher photon-shot noise. However, the noise performance in the passive image is remarkable when taking into account the 44 μs of pixel integration time and the strong ambient light attenuation. Indeed, a significant fraction of the ambient light is blocked by the optical near-infrared interference filter that is placed in front of the chip sensor (see Section II).

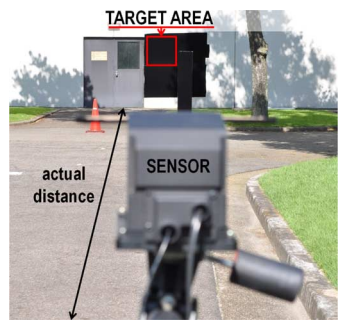

Fig. 17. Photograph of the outdoor experimental setup. The depth sensor was mounted on a tripod and directed towards the target along an internal road on our campus, capable of accommodating the required maximum distance of 100 m. It was then moved from 10 to 100 m in steps of 10 m. A tape measure was utilized to measure the actual (ground truth) distances.

C. Quantitative Evaluation and Related Discussion

In order to compare the performance of our sensor to the published state-of-the-art, we have also carried out a complete

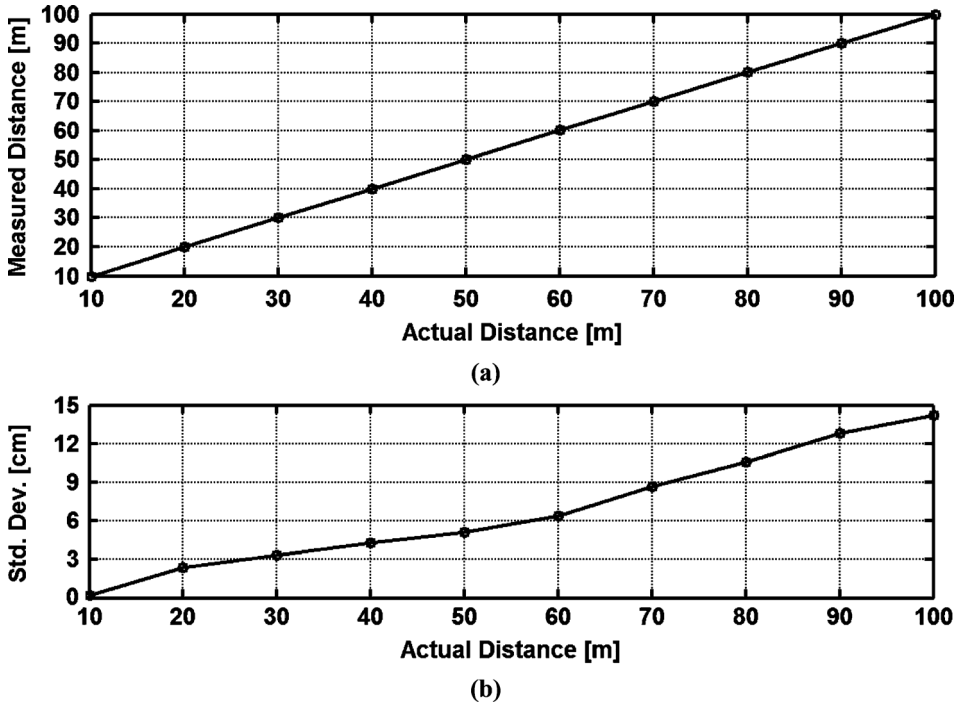

Fig. 18. (a) Mean distance values within each set of 1000 measurements as a function of the actual distance for a 9%-reflectivity noncooperative target at 10 frame/s. (b) Standard deviation as a function of the actual distance.

quantitative evaluation of the sensor system. Fig. 17 shows a photograph of our outdoor experimental setup. Similarly to previous works [21], [23], [24], the depth sensor was mounted on a tripod and directed towards the target along an internal road on our campus, which was capable of accommodating the required maximum distance of 100 m. With the exception of [21], ranging performance has usually been reported in the related literature by means of high-reflectivity and noncooperative targets, typically a flat object covered with white paper, e.g., in [5], [10]–[20], [23], and [24]. In this work, we prepared a non-cooperative target covered by a controlled 9%-reflectivity (black) material, similar to the wide board utilized in [21]. The 9%-reflectivity material was chosen to model the worst-case reflectivity potentially found on common objects. For this purpose, we utilized the external wall of one building on our campus, as shown in Fig. 17. The sensor was placed in front of the covered wall and moved from 10 to 100 m in steps of 10 m. At each distance, 1000 depth image samples were acquired at 10 frame/s. The actual target distances with respect to the sensor were controlled by means of a tape measure. A luxmeter was utilized to measure the solar background illuminance. When placing the luxmeter on the sunlight-exposed target surface, and therefore orienting it towards the LIDAR sensor, a background illuminance of 70 klux was recorded. However, when the luxmeter was oriented towards the sky under the same conditions, readings of 100 klux and above were recorded, thus demonstrating that the amount of sunlight reaching the ground was not far from the worst-case conditions. In Fig. 18(a), the mean measured distance for each set of 1000 samples is plotted as a function of the actual distance. Fig. 18(b) shows the standard deviation as a function of the actual distance within the same distance range. The biggest repeatability error for the 9%-reflectivity target was 14.2 cm throughout the distance range of 100 m. The worst-case nonlinearity error over the entire range was 11.3 cm. However, a non-negligible fraction of this error may be attributed to the uncertainty in the measurement of the ground truth positions.

A comparison of ranging performance among a number of CMOS sensors from the recently published state-of-the-art is shown in Table I. As was the case in [21], some existing TOF sensors that are commercially available, e.g., [4], [8], could not be added to the comparison table due to the unavailability of complete performance information and/or associated measurement conditions. With the exception of [21] and this work, the repeatability errors of the remaining sensors in Table I were characterized under considerably lower or undisclosed ambient light conditions. Under strong daylight conditions, our sensor achieves a relative precision, i.e., the ratio of the repeatability error to the distance range, of 0.14%. In [21], a similar precision figure was achieved by means of a high-reflectivity target (white) and by approximately twice as high as the current signal power. When imaging the same target that was utilized in this work, the measurement range in [21] was limited to 60 m, as opposed to 100 m currently. This work therefore improves the range of the previous sensor by 66%, yet utilizes approximately half of the signal power. Our proposed sensor therefore achieves superior ranging performance when compared to any of remaining sensors of Table I, even when ignoring critical parameters such as background light conditions. Considerably smaller non-linearity was also obtained with the new sensor without any postprocessing calibration. This performance was obtained with the proposed SoC approach, thus obviating for the first time, to the best of our knowledge, the need for costly companion FPGA devices such as those used in [21]. Our sensor therefore compares favorably to the state of the art, to the best of our knowledge.

TABLE I
COMPARISON OF PERFORMANCE TO RECENTLY PUBLISHED STATE-OF-THE-ART DEVICES IN CMOS

Parameter	Unit	Stoppa, *et al.*, JSSC, 2011 [17]	Walker *et al.*, ISSCC, 2011 [18]	Kim *et al.*, ISSCC, 2012 [19]	Niclass *et al.*, JSSC, 2013 [21]	Olga S. *et al.*, ISSCC, 2013 [20]	Ito *et al.*, Ph.J, 2013 [24]	**This Work**
Depth image resolution	pix.	80x60	128x96	480x270	340x96	64x64	256x64	**202x96**
CMOS process	-	0.18µm imaging	0.13µm imaging	0.11µm imaging	0.18µm HV	0.35µm 1P4M	0.18µm HV	**0.18µm HV**
Optical fill factor	%	24	3.2	34.5	70	25.7	13.9	**70**
Illumination wavelength	nm	850	850	850	870	850	870	**870**
Illumination repetition rate or modulation frequency	MHz	20	3.33	20	0.2	25	0.6	**0.133**
Illumination (signal) power	mW	80	50	not stated	40	~300[e]	75	**21**
Field of view (HxV)	deg.	7.6x5.7 [a]	40x40	24x14 [a]	170x4.5	15x15	45x11	**55x9**
Frame rate	fps	5[b]	20	11[b]	10	50	10	**10**
Target reflectivity	-	white	white	white	white	white	white	**9%**
Distance range (measured)	m	6	2.4	4.5	100	4.75	25	**100**
Unambiguous distance range	m	7.5	45	7.5	128	6	128	**128**
Background light	klux	not stated	0.11	1.5[c]	80	not stated	10	**70**
Repeatability error (1σ)	cm	16	16	3.8	<10	5.7	14.5	**14.2**
Max. relative non-linearity	%	0.7	0.21	0.93[d]	0.37	1.7	0.68	**0.11**
Relative precision	%	2.7	6.7	0.84	0.1	1.2	0.58	**0.14**

(a) Calculated using published parameters as $\text{FOV} = 2 \cdot \tan^{-1}\left(\frac{h}{2f}\right)$, where f is the lens focal length and h is the pixel array width or height. The FOV was not explicitly reported in the respective references.
(b) Calculated as the inverse of the published image acquisition time.
(c) Indoor fluorescent lighting.
(d) Pixel-level calibration applied.
(e) Only the peak optical radiance in [W/m^2] at 1m was explicitly reported. The total illumination power was approximated here by assuming a beam with circular top-hat profile and using the published FWHM FOV with a 50% modulation duty cycle.

VI. CONCLUSION

With the emerging need for high-resolution LIDAR technologies in ADASs, we have introduced a SoC that performs time-correlated single-photon counting and complete digital signal processing for a TOF sensor. At the core of the SoC, linear arrays of 16 TOF and 32 intensity-only macro-pixels based on SPADs have been utilized in an original look-ahead concept, thus acquiring active TOF and passive intensity images simultaneously. The SoC also comprises an array of circuits capable of generating precise triggers upon spatiotemporal correlation events, an array of 64 12-b TDCs, and 768 kb of on-chip SRAM memory. The SoC provides the system-level electronics with a serial and low bit-rate digital interface for: 1) multi-echo distance; 2) distance reliability; 3) intensity; and 4) passive-only intensity, thus mitigating system-level complexity and cost. The sensor chip has been fabricated in a 0.18-μm CMOS technology and has been confirmed to be fully functional. A proof-of-concept prototype achieving depth imaging up to 100 m with a resolution of 202 $\times$ 96 pixels at 10 frames/s has been introduced. A six-facet polygonal mirror rotating at 10 revolutions/s performs laser scanning with a single laser diode within a field-of-view of 55 $\times$ 9 degrees.

Quantitative evaluation of the TOF sensor under strong solar background light, i.e., 70 klux, and utilizing a low-reflectivity and noncooperative target, revealed a repeatability error of better than 15 cm throughout the distance range of 100 m, thus leading to a relative precision of 0.14%. Under the same conditions, the relative nonlinearity error, exempt from any type of calibration, was 0.11%. To the best of our knowledge, these figures compare favorably to the state of the art, in particular when taking into account critical parameters that directly impact performance such as the number of pixels, signal power, field-of-view, and frame rate.

In order to investigate the suitability of our approach in ADAS-related applications, extensive experimental data have also been obtained in normal traffic situations by having the sensor mounted on an experimental car. An example of the images that had been obtained with the sensor operating in daylight conditions has also been reported. Finally, the multi-echo and passive imaging capabilities of the proposed LIDAR have been demonstrated with a special experiment whereby the

sensor successfully imaged targets in distances up to about 80 meters through a window in our building.

ACKNOWLEDGMENT

The authors would like to thank M. Maeda for helping with the design and assembly of the optical system, M. Lany and Prof. R. S. Popovic of EPFL as well as T. Kindo of Toyota Motor Corporation for their technical support with the design of the SPAD device.

REFERENCES

[1] M. Lesemann, J. Lutzow, and A. Zlocki, "eVALUE—A test programme for active safety systems," in *Proc. 17th ITS World Congress*, Oct. 2010, pp. 1–6.

[2] 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.

[3] S. B. Gokturk, H. Yalcin, and C. Bamji, "A time-of-flight depth sensor—system description, issues and solutions," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. Workshops*, 2004, p. 35.

[4] T. Oggier, B. Buttgen, F. Lustenberger, R. Becker, B. Rüegg, and A. Hodac, ETH, Zurich, Switzerland, "SwissRanger SR3000 and first experiences based on miniaturized 3D-TOF cameras," in *Proc. 1st Range Imaging Res. Day*, 2005, pp. 97–108.

[5] C. Niclass, A. Rochas, P.-A. Besse, and E. Charbon, "Design and characterization of a CMOS 3-D image sensor based on single photon avalanche diodes," *IEEE J. Solid-State Circuits*, vol. 40, no. 9, pp. 1847–1854, Sep. 2005.

[6] O. Elkhalili, O. M. Schrey, W. Ulfig, W. Brockherde, B. J. Hosticka, P. Mengel, and L. Listl, "A 64 $\times$ 8 pixel 3-D CMOS time of flight image sensor for car safety applications," in *Proc. 32nd Eur. Solid-State Circuits Conf.*, 2006, pp. 568–571.

[7] G. Yahav, G. J. Iddan, and D. Mandelboum, "3D imaging camera for gaming application," in *Proc. IEEE Int. Conf. Consumer Electron.*, 2007, pp. 1–2.

[8] T. Ringbeck, ETH, Zurich, Switzerland, "A 3D time of flight camera for object detection," in *Proc. 8th Conf. Opt. 3D Meas. Tech.*, 2007.

[9] S. Kawahito, I. A. Halin, T. Ushinaga, T. Sawada, M. Homma, and Y. Maeda, "A CMOS time-of-flight range image sensor with gates-on-field-oxide structure," *IEEE Sensors J.*, vol. 7, no. 12, pp. 1578–1586, 2007.

[10] C. Niclass, C. Favi, T. Kluter, M. Gersbach, and E. Charbon, "A 128 $\times$ 128 single-photon imager with on-chip column level 10b time-to-digital converter array," *IEEE J. Solid-State Circuits*, vol. 43, no. 12, pp. 2977–2989, Dec. 2008.

[11] A. A. Dorrington, M. J. Cree, D. A. Carnegie, A. D. Payne, R. M. Conroy, J. P. Godbaz, and A. P. P. Jongenelen, "Video-rate or high-precision: A flexible range imaging camera," *Proc. SPIE*, vol. 6813, 2008, Art. ID 681307.

[12] B. Buttgen and P. Seitz, "Robust optical time-of-flight range imaging based on smart pixel structures," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 55, no. 6, pp. 1512–1525, Jun. 2008.

[13] W. Van der Tempel, R. Grootjans, D. Van Nieuwenhove, and M. Kuijk, "A 1 k-pixel 3D CMOS sensor," *IEEE Sensors*, pp. 1000–1003, 2008.

[14] C. Niclass, C. Favi, T. Kluter, F. Monnier, and E. Charbon, "Single-photon synchronous detection," *IEEE J. Solid-State Circuits*, vol. 44, no. 7, pp. 1977–1989, Jul. 2009.

[15] G. Zach, M. Davidovic, and H. Zimmermann, "A 16 $\times$ 16 pixel distance sensor with in-pixel circuitry that tolerates 150 klx of ambient light," *IEEE J. Solid-State Circuits*, vol. 45, no. 7, pp. 1345–1353, Jul. 2010.

[16] S. Bellisai, F. Guerrieri, and S. Tisa, "3D ranging with a high speed imaging array," in *Proc. Conf. Ph.D. Res. Microelectron. Electron.*, Jul. 18–21, 2010, pp. 1–4.

[17] D. Stoppa, N. Massari, L. Pancheri, M. Malfatti, M. Perenzoni, and L. Gonzo, "A range image sensor based on 10 μm lock-in pixels in 0.18 μm CMOS imaging technology," *IEEE J. Solid-State Circuits*, vol. 46, no. 1, pp. 248–258, Jan. 2011.

[18] R. J. Walker, J. A. Richardson, and R. K. Henderson, "A 128 $\times$ 96 pixel event-driven phase-domain $\Delta\Sigma$-based fully digital 3D camera in 0.13 μm CMOS imaging technology," in *IEEE Int. Solid-State Circuits Conf. Dig. Tech. Papers*, 2011, pp. 410–412.

[19] S.-J. Kim, B. Kang, J. D. K. Kim, K. Lee;, C.-Y. Kim, and K. Kim, "A 1920 $\times$ 1080 3.65 μm-pixel 2D/3D image sensor with split and binning pixel structure in 0.11 μm standard CMOS," in *ISSCC Dig. Tech. Papers*, Feb. 2012, pp. 396–397.

[20] O. Shcherbakova, L. Pancheri, G.-F. Dalla Betta, N. Massari, and D. Stoppa, "3D camera based on linear-mode gain-modulated avalanche photodiodes," in *ISSCC Dig. Tech. Papers*, Feb. 2013, pp. 490–491.

[21] C. Niclass, M. Soga, H. Matsubara, S. Kato, and M. Kagami, "A 100 m-range 10-frame/s 340 $\times$ 96-pixel time-of-flight depth sensor in 0.18 μm CMOS," *IEEE J. Solid-State Circuits*, vol. 48, no. 2, pp. 559–572, Feb. 2013.

[22] B. Schwarz, "LIDAR: Mapping the world in 3D," *Nature Photon.*, vol. 4, pp. 429–430, 2010.

[23] C. Niclass, K. Ito, M. Soga, H. Matsubara, I. Aoyagi, S. Kato, and M. Kagami, "Design and characterization of a 256 $\times$ 64-pixel single-photon imager in CMOS for a MEMS-based laser scanning time-of-flight sensor," *Opt. Exp.*, vol. 20, pp. 11863–1188, 2012.

[24] K. Ito, C. Niclass, I. Aoyagi, H. Matsubara, M. Soga, S. Kato, M. Maeda, and M. Kagami, "System design and performance characterization of a MEMS-based laser scanning time-of-flight sensor based on a 256 $\times$ 64-pixel single-photon imager," *IEEE Photon. J.*, vol. 5, no. 2, 2013, Art. ID 6800114.

[25] S. Cova, A. Longoni, and A. Andreoni, "Towards picosecond resolution with single photon avalanche diodes," *Rev. Sci. Instrum.*, vol. 52, pp. 408–412, 1981.

[26] C. Niclass, M. Soga, H. Matsubara, M. Ogawa, and M. Kagami, "A 0.18 μm CMOS SoC for a 100 m-range 10-frame/s 200 $\times$ 96-pixel time-of-flight depth sensor," in *ISSCC Dig. Tech. Papers*, Feb. 2013, pp. 488–489.

[27] C. Niclass, M. Soga, H. Matsubara, and S. Kato, "A 100 m-range 10-frame/s 340 $\times$ 96-pixel time-of-flight depth sensor in 0.18 μm CMOS," in *Proc. IEEE Eur. Solid-State Circuits Conf.*, 2011, pp. 107–110.

Cristiano Niclass (S'05–M'08) received the M.S. degree in microelectronics and Ph.D. degree in electrical engineering and computer science from the Swiss Federal Institute of Technology, Lausanne, Switzerland, in 2003 and 2008, respectively.

From 2003 to 2006, he was a part-time R&D Engineer with Ingenico (Suisse) SA, working on the design of electronic payment devices. In 2003, he joined the Processor Architecture Laboratory, Swiss Federal Institute of Technology Lausanne (EPFL), Lausanne, Switzerland. At EPFL, he has investigated the design, implementation, and characterization of single-photon image sensors in CMOS. He has also worked on the design of high-speed and high-resolution data converters. In 2009, he joined the Applied Optics Lab, Toyota Central R&D Labs Inc., Nagakute, Japan, where he has been working in the field of CMOS sensors. He has authored and coauthored over 60 peer-reviewed journal and conference publications. He holds seven patents and is the inventor or a co-inventor in a number of recent patent applications. His current interests include many aspects of high-speed and low-noise mixed-signal ICs, with emphasis on high-performance imaging.

Mineki Soga received the B.E. degree in electrical and electronic engineering and M.E. degree in information processing from the Tokyo Institute of Technology, Tokyo, Japan, in 1993 and 1995, respectively.

Since then, he has been with the Toyota Central R&D Labs Inc. (TCRDL), Nagakute, Japan. At TCRDL, he has worked on image sensing and computer vision, such as pedestrian detection for night view systems. From 2005 to 2007, he was dispatched to Toyota Motor Europe NV/SA (TME) Belgium. At TME, he investigated laser rangefinder using CMOS single-photon detectors in collaboration with the Swiss Federal Institute of Technology Lausanne. His current interests include sensing and recognition of road environment for automotive safety.

Hiroyuki Matsubara received the B.E. degree in applied physics and M.E. degree in optical engineering from Hokkaido University, Sapporo, Japan, in 1988 and 1990, respectively.

Since 1990, he has been with the Toyota Central Research and Development Laboratories Inc., Nagakute, Japan, where he is currently a Senior Researcher. His current research interests include several aspects of optical measurement systems with emphasis on their model-based optical design.

Mr. Matsubara is a member of the International Society for Optical Engineering (SPIE).

Masaru Ogawa received the B.S., M.S., and D.E. degrees from the Nagoya Institute of Technology, Nagoya, Japan, in 1990, 1992, and 2010, respectively, all in electrical engineering and computer science.

From 1992, he has been with the Toyota Central R&D Labs Inc. (TCRDL), Nagakute, Japan. At TCRDL, he has been engaged in research on vehicle antennas for satellite communications, automotive millimeter-wave radars and laser radar systems. He is now a Senior Researcher with the Environment Recognition Sensor Laboratory.

Dr. Ogawa is a member of the Institute of Electronics, Information and Communication Engineers (IEICE).

Manabu Kagami (M'04) received the B.E., M.E., and D.E. degrees in information engineering from the Toyohashi University of Technology, Toyohashi, Japan, in 1982, 1984, and 1999, respectively.

In 1984, he joined Mitsubishi Rayon Corporation, where he worked on fiber components and crystal growth for light-emitting devices as a Research Engineer. In 1992, he joined Toyota Central R&D Laboratories, Inc., Aichi, Japan, where he is a Principal Researcher with the Information and Communication Research Division. He is currently also serving as a Visiting Professor with the Toyota Technological Institute. His research interest includes the design of polymeric integrated optical circuits for interconnections and communication modules, optical network systems for automotive applications, and optical sensing device and systems.

Dr. Kagami is a senior member of the Optical Society of America and the Institute of Electronics, Information and Communication Engineers (IEICE). He has contributed to the development and standardization of test methods for multimode optical fiber network devices at IEC/TC86/SC86B as an expert member. He was the recipient of a Best Paper Award from IEICE in 2008.