

R49 Image Sensor Capable of Analog Convolution for Real-time Image Recognition System Using Crystalline Oxide Semiconductor FET

Seiichi Yoneda, Yusuke Negoro, Hidetomo Kobayashi, Kosei Nei, Toshihiko Takeuchi, Masashi Oota, Takuya Kawata, Takayuki Ikeda, and Shunpei Yamazaki
Semiconductor Energy Laboratory, Hase, Atsugi-shi, Kanagawa 243-0036, Japan
Phone: +81-46-248-1131 Fax: +81-46-270-3751, E-mail: sy0936@sel.co.jp

Abstract

To build a low-power real-time image recognition system, an image sensor capable of convolution of a captured image and filter data in pixels is prototyped, and its operation is demonstrated. This image sensor uses crystalline oxide semiconductor (OS) FETs for all transistors and is constructed with an analog multiply accumulator utilizing low off-state current of the OSFETs. Contour extraction of the captured image is confirmed by the demonstration.

Introduction

In image recognition, deep learning has been adopted as a convolution neural network (CNN) algorithm [1]. CNN includes processing of convolution where multiply-accumulate (MAC) operation is conducted on imaging data and filter data while the filter data is shifted. This operation has been done in a digital manner in GPU. In an image recognition system demanding timeliness, such as a dashboard camera, high-speed MAC operation on imaging data captured by an image sensor is required, and thus a load on GPU is increased, resulting in large power consumption of the whole system.

The research on oxide semiconductors (OSs) started from the synthesis of indium-gallium-zinc oxide (IGZO) by Dr. Kimizuka et al [2–6]. Figure 1 shows classification of IGZO crystal structures. We have discovered c-axis aligned crystalline IGZO (CAAC-IGZO) [7], nano crystalline IGZO (nc-IGZO) [8–10], and cloud-aligned composite IGZO (CAC-IGZO) [11]. According to Dr. Kimizuka's opinion, the IGZO morphology we have discovered belongs to an "intermediate state" between an amorphous structure and a crystal structure [12], and thus the morphology classified as crystalline IGZO in Fig. 1. In particular, CAAC-IGZO brings a feature of low off-state current of FETs where CAAC-IGZO is used for an active layer (OSFETs) [17]. Utilizing this feature, a global shutter image sensor using OSFETs in pixels [18–20] and a memory-integrated analog multiply accumulator with high power efficiency [21] have been reported. By a combination of both concepts of the previous reports, it is possible to conduct low-power convolution of filter data and captured imaging data while reading out is conducted. Such an image sensor enables a reduction in load on GPU, resulting in achievement of a low-power real-time image recognition system. This study reports an image sensor capable of analog convolution for the low-power real-time image recognition system, and its operation verification.

Circuit Configurations

Figure 2 is a block diagram of the image sensor in this study. This image sensor adopts OSFETs for all transistors and a global shutter system which enables images of a moving object at high speed to be captured without distortion. In other words,

our image sensor is advantageous in image recognition.

A principle of analog MAC operation in the image sensor is explained with reference to Fig. 3. When select transistors in pixels are turned on, read transistors satisfy saturation conditions, and the drain current value of the read transistors is defined as $I_d = \beta(V_{gs} - V_{th})^2/2$. Transistors in an I-V converter each have a constant resistant value R without depending on the voltage of a read line WX. The voltage change amount at a node FD in a pixel is denoted by X_i , which is generated by transfer of photocharge of a photodiode. The voltage of filter data supplied from a signal line W[8:0] is denoted by W_i .

After the FD in each pixel is reset, read voltages corresponding to two conditions are obtained as shown in Fig. 3(b). One of the conditions is to supply filter data W_i , and the other is to supply blank filter data (0 V filter data). The difference of the two read voltages is generated in a correlated double sampling (CDS) circuit to be a voltage V_1 . When the same processing is conducted after photocharge is transferred to the FD which has been reset as shown in Fig. 3(c), a voltage V_2 is obtained. The difference of V_1 and V_2 , which is denoted by $V_2 - V_1 = \sum_i \beta X_i W_i R/3$, is calculated outside the circuit. In the above manner, imaging data and filter data are processed in MAC operation.

Figure 4 is a timing chart used to describe a process of convolution of the imaging data and the filter data with the above analog MAC operation conducted by the image sensor shown in Fig. 2. After the FD in all pixels are reset, units (one unit is composed of 3×3 pixels) subjected to MAC operation are selected with a row driver and switch control signals (SY). The row driver activates three adjacent selection signal lines at once, and SY short-circuits three read lines WX in columns adjacent to each other. By this, 80 units of 3×3 pixels are selected at once, and voltages that are output from the selected 80 units (80 types of voltage) are input to the CDS circuit. At this stage, the following operation is conducted: the filter data W_i is supplied, the CDS circuit is reset, and subsequently the blank filter data is supplied. By this, 80 types of voltages comparable to V_1 are generated in the CDS circuit. The unit (3×3 pixels) selection by the row driver and SY is shifted in sequence, and accordingly 80 types of V_1 are read out to the outside in sequence. Next, the pixels are reset, and an image is captured. Then, the above-described process (from the unit selection to voltage read out to the outside) is conducted in a similar manner, so that voltages comparable to V_2 are read out to the outside. After that, the difference between V_1 and V_2 is calculated outside, resulting in the MAC operation, that is, convolution, on all the shifted combinations.

The image sensor in this study also conducts normal imaging operation. For the normal operation, the voltages of VIV and VBIAS are controlled to make the transistors in the I-V converter serve as bias transistors of source followers, and the

row driver is driven to activate the selection signal lines sequentially row by row. Such a configuration enables a general image sensor to have a convolution operation without mounting additional devices. Thus, there is no necessity to assign an area to the convolution operation.

Fabrication and Measurement

A process of OSFET with 0.5- μm technology (Fig. 5) was used to prototype the image sensor (see Fig. 6 and Table 1). Crystalline selenium, which is compatible with the OSFET process, was used for a photoelectric conversion film [22–23] (Fig. 7).

The photocharge-induced voltage change amount X at FD was simulated by a change of a reset voltage V_{RS} of pixels, and filter data W was swept with respect to multiple X values, so that multiplication performance was measured (Fig. 8). One unit of 3×3 pixels was evaluated in this measurement, and the voltage values of X and W were the same in all pixels. The multiplication performance with 4-bit accuracy was observed within a range of $X \leq 0.5$ V. Next, imaging and convolution with two types of filter data were conducted to extract a feature value of the image (Fig. 9). By convolution with filter data of horizontal stripes and vertical stripes, the horizontal-direction component and the vertical-direction component in a contour of the zebra image were extracted. Moreover, the image filled in black and white was rotated around one straight line as a boundary to evaluate feature extraction performance by horizontal stripe filters. A rotation center was positioned in the evaluated unit of 3×3 pixels as shown in Fig. 10(b). The difference of MAC operation results obtained from the rotated image was defined as a feature value. According to the relation between the feature value and the rotation angle as shown in Fig. 10(a), the feature was extracted most clearly at 0° and the feature extraction was confirmed up to approx. 40° . In other words, convolution was demonstrated in the image sensor in this study.

Conclusions

This study has demonstrated that convolution of a captured image and filter data is possible in our prototyped image sensor using OSFETs with low off-state current. Application of this image sensor is expected to achieve a low-power real-time image recognition system.

References

[1] H. Jeong et al., International Conference on Computational Science and Computational Intelligence, pp. 824–828, 2016.
 [2] Japanese Published Patent No. S63-239117.
 [3] N. Kimizuka and T. Mohri, J. Solid State Chem., 60, pp.382-384 (1985).
 [4] M. Nakamura et al., J. Solid State Chem., 93, pp.298-315 (1991).
 [5] N. Kimizuka et al., J. Solid State Chem., 116, pp.170-178 (1995).
 [6] C. Li et al., J. Solid State Chem., 139, 347–355 (1998). [10] S. Yamazaki et al., Journal of SID., 22, No. 1, pp. 55-67 (2014).
 [7] S. Yamazaki et al., Journal of SID., 22, No. 1, pp. 55-67 (2014).
 [8] S. Ito et al., Proc. AM-FPD'13 Digest, pp. 151-154 (2013).
 [9] Japanese Published Patent No. 5894694.
 [10] US Patent 9,153,650.
 [11] S. Yamazaki et al., Jpn. J. Appl. Phys., 55, 115504 (2016).
 [12] E-mail that N. Kimizuka sent to M. Takahashi et al., on Sep. 8, 2012.
 [13] K. Nomura et al., Nature, 432, pp. 488-492 (2004).

[14] T. Kamiya et al., Proc. IDW'13 Digest, pp. 280-281 (2013).
 [15] Y. Yamada et al., Jpn. J. Appl. Phys., 53, 091102 (2014).
 [16] Y. Waseda et al., Materials Transactions, 59, No. 11, pp. 1691-1700 (2018).
 [17] K. Kato et al., Jpn. J. Appl. Phys., 51, 021201 (2012).
 [18] T. Aoki et al., Symposium on VLSI Technology, pp. 174–175, 2011.
 [19] S. Yoneda et al., Extended Abstract of Solid-State Devices and Materials, pp. 988–989, 2014.
 [20] T. Ohmaru et al., International Solid State Circuit Conference, pp.118–119, 2015.
 [21] T. Aoki et al., Extended Abstracts of Solid State Devices and Materials, pp.191–192, 2017.
 [22] Y. Kurokawa et al., International Image Sensor Workshop, 2015.
 [23] S. Imura et al., International Electron Devices Meeting, pp. 88–91, 2014.

Table 1. Specifications of prototyped image sensor

Process	0.5 μm OSFET
Die size	8.0 mm x 8.0 mm
Number of pixels	240(H) x 162(V)
Pixel size	15 μm x 15 μm
Pixel configuration	4 transistors, 1capacitor
Output	16ch analog voltage
Conversion gain	1.66 $\mu\text{V}/\text{h}+$
Full well capacity	612 kh+
Read noise	751 h+rms
Fill factor	79.8 %
Frame rate	7.5 fps
Power consumption	normal capturing: 3.80 mW convolution: 7.06 mW
Multiplication efficiency	0.805 GOP/s/W
CNN filter size	3(H) x 3(V)
CNN stride	1

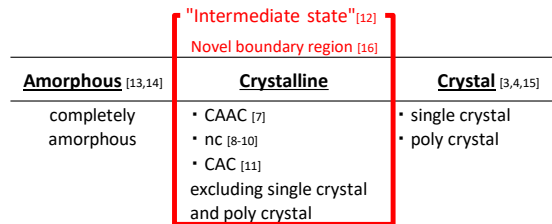


Fig. 1. IGZO classification based on crystal structure

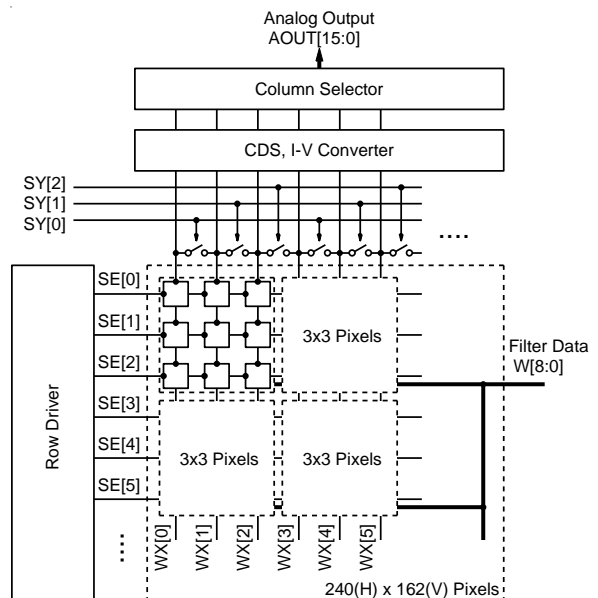


Fig. 2. Block diagram of image sensor

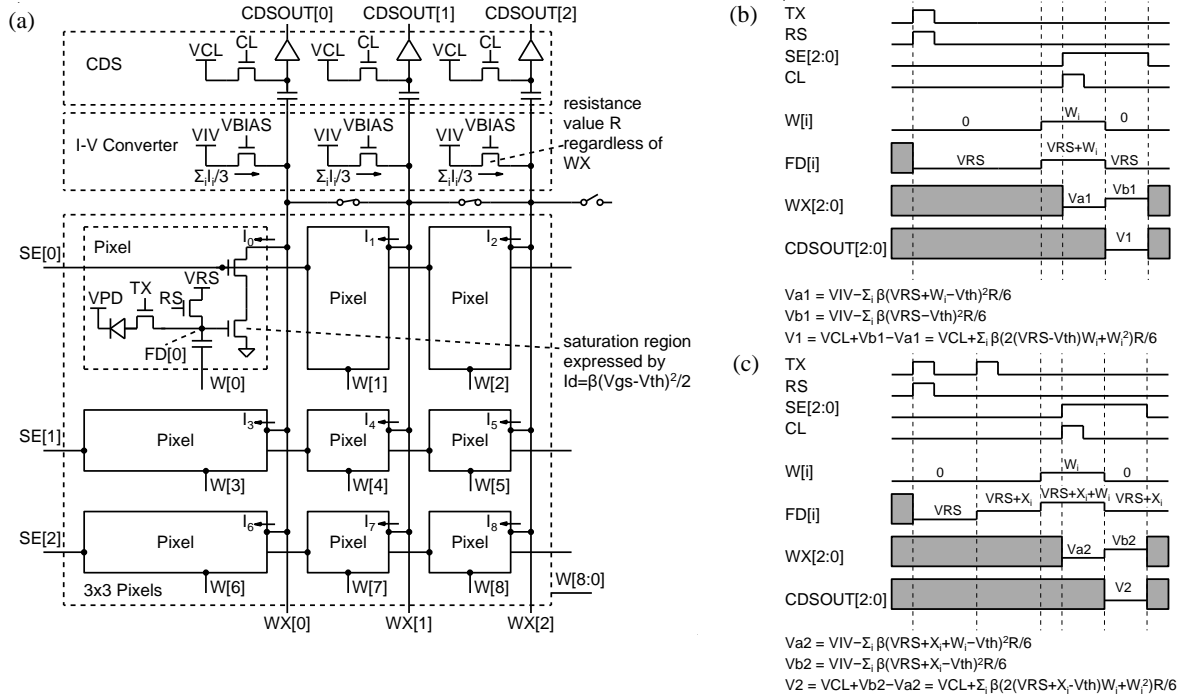


Fig. 3. Pixel unit (3x3 pixels) and peripheral circuits: (a) circuit diagram, (b) and (c) timing chart

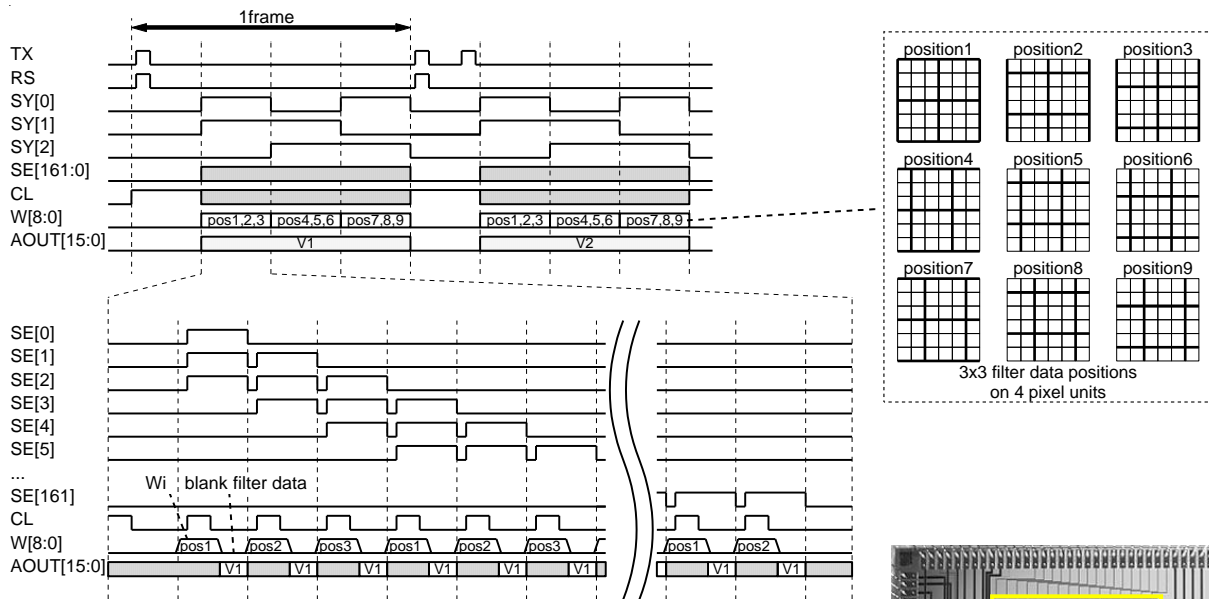


Fig. 4. Timing chart of image sensor

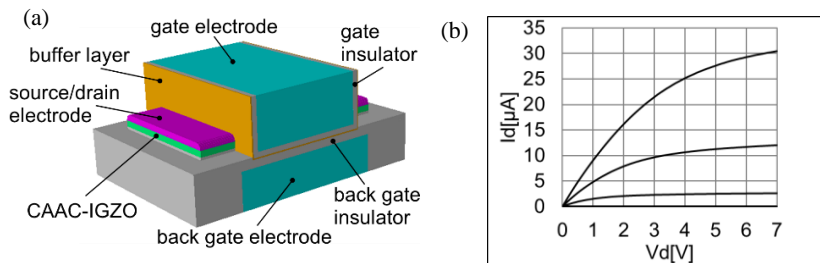


Fig. 5. (a) Perspective view of OSFET ($W/L = 0.5\mu\text{m}/0.5\mu\text{m}$)
 (b) V_d - I_d performance of OSFET ($V_g = 1, 3, 5, 7V$) ($W/L = 0.5\mu\text{m}/0.5\mu\text{m}$)

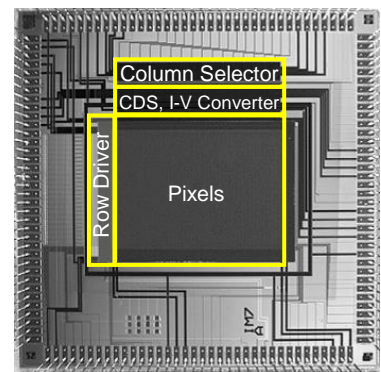


Fig. 6. Picture of prototyped image sensor

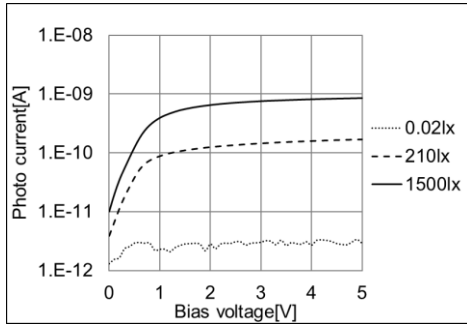


Fig. 7. Photoelectric current performance of crystalline selenium

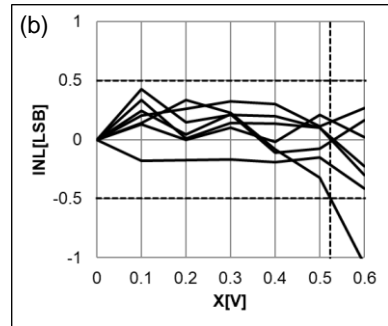
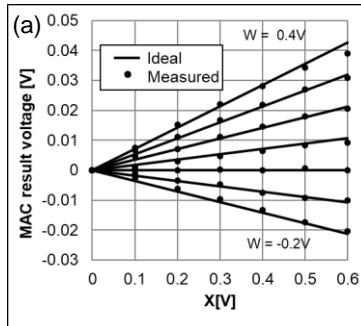


Fig. 8. Multiplication performance of image sensor: (a) multiplication results of X and W, (b) integrated nonlinearity error (4 bit)

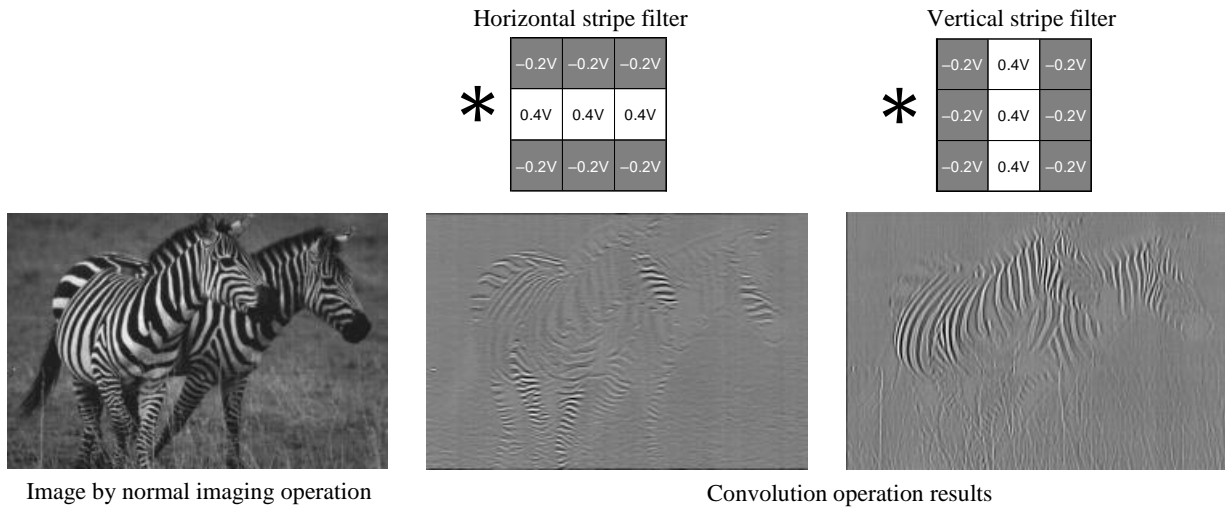


Fig. 9. Image feature extraction

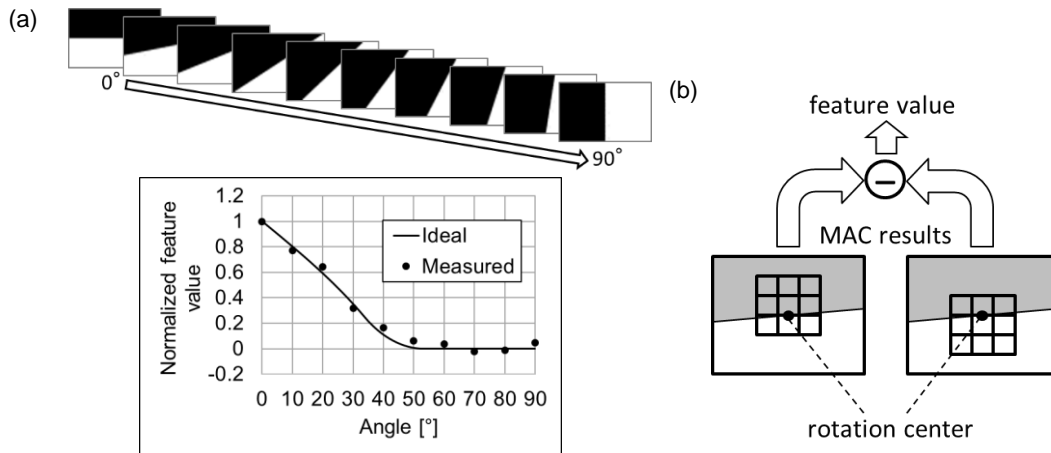


Fig. 10. (a) Feature extraction performance, (b) Calculation of feature value