Extended Abstracts of the 2019 International Conference on Solid State Devices and Materials, Nagoya, 2019, pp385-386

Crystalline Oxide Semiconductor FET/Si-FET Hybrid Structure-based In-Memory Computing Circuit for Artificial Neural Network Applications

Takeshi Aoki, Munehiro Kozuma, Yoshiyuki Kurokawa, Hajime Kimura and Shunpei Yamazaki

Semiconductor Energy Laboratory Co., Ltd.
398, Hase, Atsugi-shi, Kanagawa 243-0036, Japan
Phone: +81-46-270-1170 E-mail: ta1107@sel.co.jp

Abstract

We examined the implementability of an in-memory computing circuit for artificial neural network using programmable current source circuit and a memory formed of an oxide semiconductor (OS) FET including crystalline OS in its channel layer.

1. Introduction

Energy reduction in memory access and data transfer is an issue for an artificial neural network (ANN) using the von Neumann computer [1]. To solve the issue, in-memory computing (IMC) in which arithmetic operation is also performed in a memory circuit in data reading has attracted attention. Extremely low off-state current of an oxide semiconductor FET (OS-FET) realizes a charge retention multilevel memory (OS memory) [2, 3], while it is difficult for an SRAM to retain multilevel data [1]. In addition, unlike to a flash memory [5], voltage to be retained can be directly input to the OS memory like a DRAM [2]. Furthermore, unlike an ReRAM [6], the OS memory does not retain data in a memory element, and therefore is robust against process variation (Table I). In view of this, we examined an OS memory-based IMC circuit for ANN.

2. Circuit Design with OS Multi-bit Memory for IMC

Circuit : OS Memory Cell, Current Source and IMC Circuit

Fig. 1 shows circuit operation of the OS memory, which consists of one OS-FET, one Si-FET, and one capacitor. In data writing, weight data W is written to a node FN ($V_{FN} = \Delta W + V_{ref}$). In data reading, input data X is written to the node FN by capacitive coupling ($V_{FN} = \Delta W + V_{ref} + \Delta X$). In addition, the IMC circuit utilizes current in a saturation region of the Si-FET. As expressed by the equation (1) in Fig. 1, the product of $\Delta W \times \Delta X$ is obtained by addition and subtraction of currents I_1 to I_4 of the Si-FETs with respect to four kinds of V_{FN} [4].

Fig. 2 illustrates a configuration of a circuit capable of efficiently generating the four currents I_1 to I_4. Two memory cells, i.e., *Cell A* to which ΔW is written and *Cell B* to which ΔW is not written are used. In data reading, currents to flow in the Si-FETs are generated in two modes: *Mode1* in which ΔX is not applied and *Mode2* in which ΔX is applied. Current I_5 is obtained by addition and subtraction of the currents I_1 to I_4 generated in the cells in the corresponding modes. Since *Mode2* operation is performed after *Mode1* operation, current source circuits for retaining the currents I_3 and I_4 generated in the *Mode1* operation are required.

Fig. 3 illustrates a programmable current source circuit using OS-FETs. In *Mode1*, the current $I_3(I_4)$ is generated in the *Cell A*(*B*) and copied to a current source *A*(*B*). In *Mode2*, the current $I_1(I_2)$ is generated in the *Cell A(B)* and a switch *A*(*B*) of OS-FET is turned off. Consequently, the current $I_3(I_4)$ retained in the current source *A*(*B*) in *Mode1* flows, and the four currents I_1 to I_4 are added and subtracted to give the current I_5 in proportion to $\Delta W \times \Delta X$.

Fig. 4 is a diagram of the whole circuit configuration. The current I_5 is converted into voltage by an operational amplifier. Connection of numerous *Cells A* and *Cells B* in the column direction allows the product-sum operation. A plurality of *Cells A* in one row can share one *Cell B,* which is independent of the weight ΔW.

Simulation Result of the IMC Circuit

Fig. 5 shows simulation results of an IMC circuit designed on the assumption of an OS-FET based on a 350-nm technology and a Si-FET based on a 110-nm technology. Fig. 5(a) is the result of the product-sum operation $Y = \sum_{i=1}^{n=25} \Delta W_i \times \Delta X_i$ where $\Delta W_{1-25} = +1$ or -1 and $\Delta X_{1-25} = +1$, 0, or -1, which demonstrates that the product-sum operation is possible. Fig. 5(b) is the result of the product-sum operation $Y = \sum_{i=1}^{n} \Delta W_i \times \Delta X_i$ where $\Delta W_{1-n} = +1$, -1, or 0 and $\Delta X_{1-n} = +1$ or 0. We confirmed that the result of product-sum operation changes in proportion to the number of rows n. Fig. 5(c) shows the simulation results of the multiplication $\Delta W_1 \times \Delta X_1$ where $\Delta W_1 = +1$ or -1 and $\Delta X_1 = +1$ or -1 on the assumption of the Si-FET characteristic variation in the manufacturing process (3σ /1024runs). The results show that 3σ is less than 0.1 in the multiplication $\pm1 \times \pm1$, which reveals that a margin that allows a multiplication is ensured.

3. Conclusions

We designed an IMC circuit based on an OS-FET/Si-FET hybrid structure, and verified its circuit operation. This circuit is applicable to ANN applications.

References

[1] A. Biswas *et al.*, ISSCC2018, pp. 488-490.
[2] T. Matsuzaki *et al.*, IMW2015, pp. 125-128.
[3] Y. Waseda *et al.*, Mater. Trans. **59** (11), pp. 1691-1700 (2018).
[4] A. Aslam-Siddiqi *et al.*, JSSC **33** (10), pp. 1502-1509 (1998).
[5] X. Guo *et al.*, IEDM2017, pp. 151-154.
[6] W. H. Chen *et al.*, ISSCC2018, pp. 494-496.

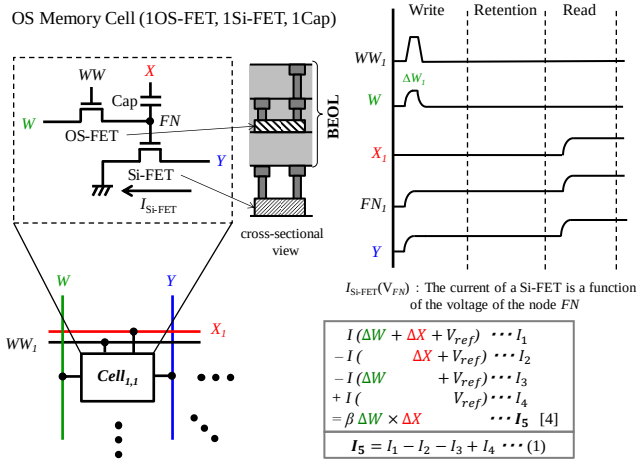

Fig. 1 Circuit operation of OS memory cell and multiplication

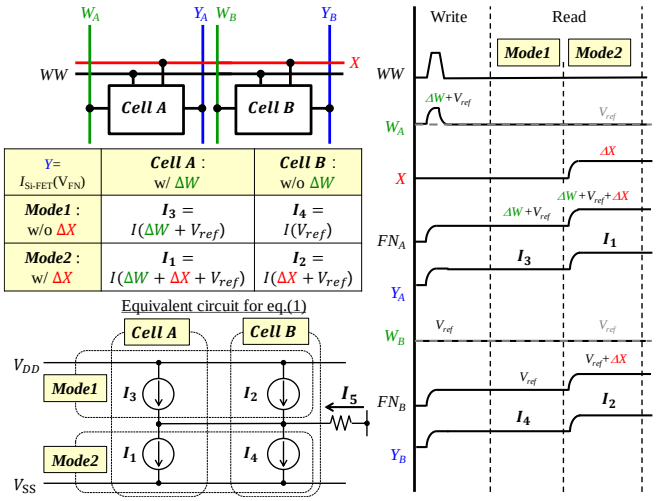

Fig. 2 Circuit operation of *Cell A and Cell B* generating I_1–I_4 and I_5 generation equivalent circuit

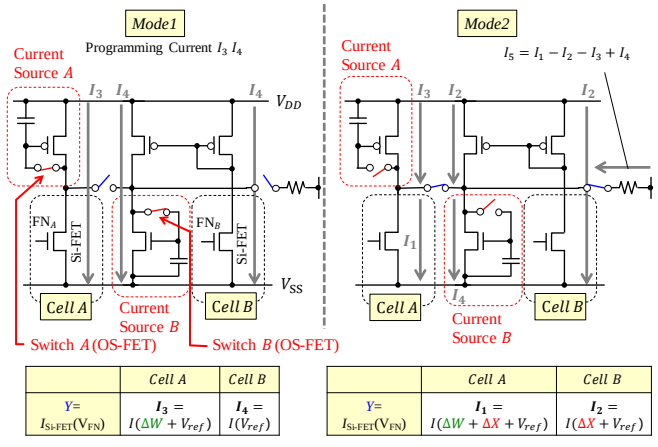

Fig. 3 Current source circuits and data reading circuit operation in *Mode1* and *Mode2*

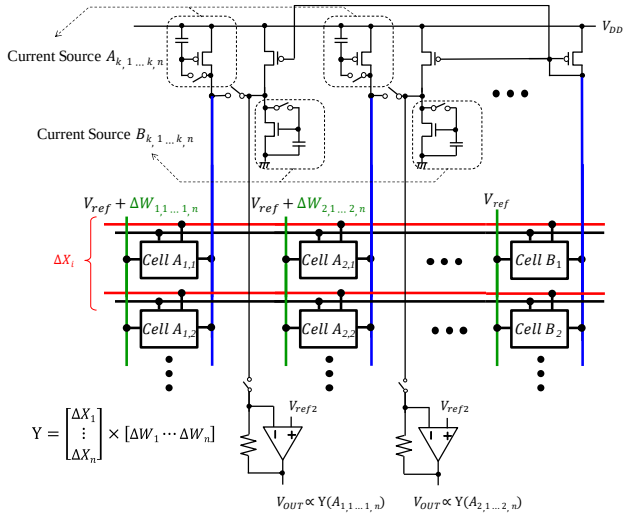

Fig. 4 Whole in-memory computing circuit

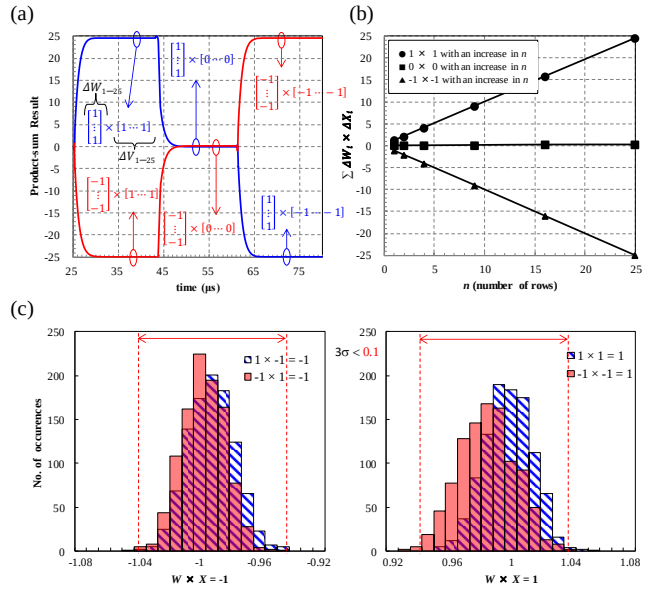

Fig. 5 Simulation results of product-sum circuit operation
(a) $\sum_{i=1}^{n=25} \Delta W_i \times \Delta X_i$: transient analysis result
(b) $\sum_{i=1}^{n} \Delta W_i \times \Delta X_i$: with an increase of n
(c) $\Delta W_1 \times \Delta X_1$: on the assumption of characteristic variation (3σ /1024runs)

Table I Comparison among memory cells for in-memory computing

	OS memory [2]	SRAM [1]	Flash [5]	ReRAM [6]
Data type	Charge Control	Inverter loop	Charge Injection	Resistance
Multi-bit vs Variation	☺	N/A	☹	☹