

Received October 13, 2020, accepted December 15, 2020, date of publication December 28, 2020, date of current version January 6, 2021.

Digital Object Identifier 10.1109/ACCESS.2020.3047619

# An Energy-Efficient Time-Domain Analog CMOS BinaryConnect Neural Network Processor Based on a Pulse-Width Modulation Approach

MASATOSHI YAMAGUCHI<sup>1</sup>, GOKI IWAMOTO<sup>1</sup>, YUTA NISHIMURA<sup>1</sup>,  
HAKARU TAMUKOH<sup>1,2</sup>, (Member, IEEE), AND TAKASHI MORIE<sup>1,2</sup>, (Member, IEEE)

<sup>1</sup>Graduate School of Life Science and Systems Engineering, Kyushu Institute of Technology, Kitakyushu 808-0196, Japan

<sup>2</sup>Research Center for Neuromorphic AI Hardware, Kyushu Institute of Technology, Kitakyushu 808-0196, Japan

Corresponding author: Takashi Morie (morie@brain.kyutech.ac.jp)

This work was supported in part by JSPS KAKENHI under Grant 22240022, Grant 15H01706, and Grant 20H04258.

**ABSTRACT** This paper proposes a time-domain analog calculations model based on a pulse-width modulation (PWM) approach for neural network calculations including weighted-sum or multiply-and-accumulate calculation and rectified-linear unit operation. We also propose very-large-scale integration (VLSI) circuits to implement the proposed model. Unlike the conventional analog voltage or current mode circuits, our circuits use transient operation in charging/discharging processes to capacitors through resistors. Since the circuits calculate multiple weighted-sums by charging a capacitance, they can be operated with extremely low energy consumption. However, because a relatively long time constant is required to guarantee calculation resolution in the time domain, they have to use very high-resistance devices, on the order of giga-ohms. We designed, fabricated, and tested a proof-of-concept complementary metal-oxide-semiconductor (CMOS) VLSI chip using a 250-nm fabrication technology to verify weighted-sum operation based on the proposed model with binary weights and PWM input signals, which realizes the BinaryConnect model. In the chip, memory cells of static-random-access memory (SRAM) are used for synaptic connection weights. High-resistance operation was realized by using the subthreshold operation region of MOS transistors, unlike in the ordinary in-memory-computing circuits. We evaluated the energy efficiency and temperature characteristics by measurement using the fabricated chip, where the highest energy efficiency for the weighted-sum calculation was 300 TOPS/W (Tera-Operations Per Second per Watt). The effects by a temperature change can be compensated for by adjusting the bias voltage. If state-of-the-art VLSI technology is used to implement the proposed model, an energy efficiency of more than 1,000 TOPS/W will be possible.

**INDEX TERMS** Artificial intelligence hardware, AI processor, deep neural networks, in-memory computing, multiply-and-accumulate, pulse-width modulation, time-domain analog computing, weighted sum, very large-scale integration (VLSI).

## I. INTRODUCTION

Artificial neural networks (ANNs), such as convolutional neural networks (CNNs) [1] and multi-layer perceptrons (MLPs) [2], have shown excellent performance on various tasks, including image recognition [2]–[6]. However, computation in ANNs is very heavy, which leads to high power consumption in current digital computers and even in highly parallel coprocessors such as graphics processing units (GPUs). In order to implement ANNs in edge devices such as

mobile phones and personal service robots, operation at very low power consumption is required.

In ANN models, weighted summation, or multiply-and-accumulate (MAC) operation, is an essential and heavy calculation task, and dedicated complementary metal-oxide-semiconductor (CMOS) very-large-scale integration (VLSI) processors have been developed to accomplish it [7]–[10]. As an implementation approach other than digital processors, the use of analog operation in CMOS VLSI circuits is a promising method for achieving extremely low-power consumption for such calculation tasks [11]–[14]. In particular, in-memory computing (IMC) approaches, which achieve weighted-sum calculation utilizing the memory circuit, such

The associate editor coordinating the review of this manuscript and approving it for publication was Yongqiang Cheng<sup>1</sup>.

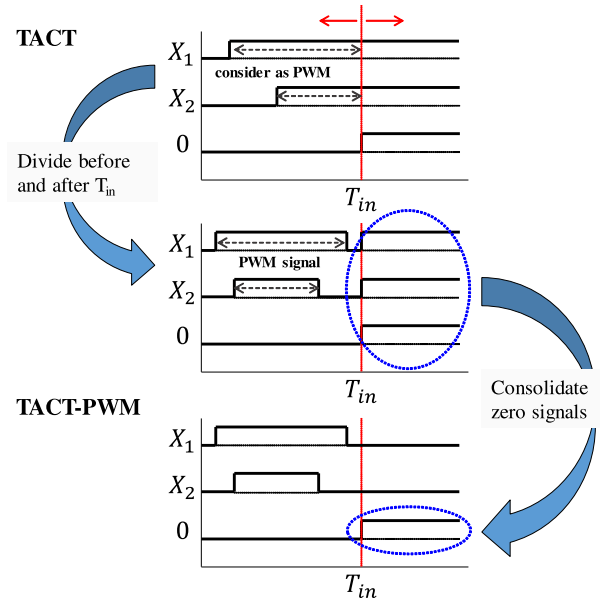
as static-random-access memory (SRAM), have been popular since around 2016 [15]–[18].

Although the calculation precision is limited due to the non-idealities of analog operation, such as noise and device mismatches, neural network models and circuits can be designed to be robust to such non-idealities [19]–[21]. On the other hand, ANN models with binarized weights called *BinaryConnect* or even with binarized inputs have been proposed as *binarized networks* and their comparable performance has been demonstrated, mainly in applications of image recognition [22], [23]. These models facilitate the development of energy-efficient hardware implementations [13]. Although the state-of-the-art AI chips with multi-bit weights and multi-bit inputs achieved an energy efficiency of up to 10 tera-operations per second per watt (TOPS/W) [24], that for the BinaryConnect model achieved up to about 50 TOPS/W [25], and those for binarized networks achieved up to 700 TOPS/W with the aid of an analog computing circuit approach [26].

The time-domain analog weighted-sum calculation model was originally proposed based on mathematical spiking neuron models inspired by biological neuron behavior [27], [28]. We have simplified this calculation model under the assumption of operation in analog circuits with transient states, and call its VLSI implementation approach “Time-domain Analog Computing with Transient states (TACT).” In contrast to conventional weighted-sum operation in analog voltage or current modes, the TACT approach is suitable for operation with much lower power consumption in the CMOS VLSI implementation of ANNs.

We have already proposed a device and circuit that perform time-domain weighted-sum calculation [29]–[31]. Some AI processors based on different time-domain approaches have been reported recently [32]–[35]. Unlike these other approaches, our proposed circuit consists of plural input resistive elements and a capacitor (RC circuit), where the weights of the network are expressed by the resistance values. The energy consumption could be lowered to the order of 1 fJ per operation, which corresponds to 1,000 TOPS/W. We also proposed a circuit architecture to implement a weighted-sum calculation with differently signed weights with two sets of RC circuits, one of which calculates positively weighted sums while the other calculates negatively weighted sums [36], [37]. Using a similar time-domain approach, a vector-by-matrix multiplier using flash memory technology was proposed [38]. However, weighted-sum calculation circuits using pulse-width modulation (PWM) signals have previously been proposed [39].

In this paper, we reformulate the weighted-sum calculation model based on the time-domain analog computing approach using PWM signals, an approach called the TACT-PWM, and propose and demonstrate its applications to ANNs such as MLPs and CNNs with extremely high computing energy efficiency. We also show the design and measurement results of a VLSI ANN chip fabricated using a 250-nm CMOS VLSI technology; we compare the calculation results by the proposed model with the ordinary numerical calculation results



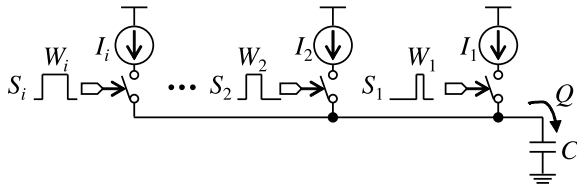
**FIGURE 1.** Modification from the TACT to TACT-PWM approach, where the label ‘0’ means an input pulse corresponding to the zero value.

and verify the very high computing efficiency of the proposed model [40]. We show the temperature characteristics measured using the fabricated chip, and suggest that the effects by a temperature change can be compensated for by adjusting the bias voltage.

## II. TIME-DOMAIN WEIGHTED-SUM CALCULATION CIRCUIT MODEL WITH PWM SIGNALS

The TACT-PWM approach can be considered a modified version of the original TACT approach [36], [37]. The modification from the TACT to the TACT-PWM is shown in Fig. 1. The original TACT approach is based on temporal coding of the integrate-and-fire neuron model, and the input/output information is expressed by the timing of signals, where inputs are given by step signals. When the weighted sum of transient responses of all input step signals exceeds the threshold, an output step signal is generated, which can be fed into the connected neurons as an input signal. According to the temporal coding in spiking neuron models, the time spans during which input and output signals are given, such as  $T_{in}$  and  $T_{out}$ , do not necessarily have to be defined, but it is useful to define such time spans for MAC calculations in CNNs.

In the TACT-PWM approach, the inputs to a neuron are expressed by PWM signals given in the input period defined by  $T_{in}$ . The weighted sum of the input PWM signals is temporarily stored, and then an output PWM signal that has a pulse width proportional to the weighted sum of the inputs is generated in the output period defined by  $T_{out}$ . The modification from the TACT to the TACT-PWM can be considered as a separation of the input step signals in the TACT into two signals in  $T_{in}$  and  $T_{out}$ , respectively, as shown in Fig. 1. Because of this modification, we can perform input and output operations independently, and do not necessarily have to



**FIGURE 2.** Weighted-sum calculation using current sources switched with PWM signals.

set  $T_{in}$  and  $T_{out}$  consecutively, which is useful for shortcut connections in ResNets [41]. In addition, no pulses have to be given for zero-value inputs as in the TACT approach.

The basic circuit configuration based on the TACT-PWM approach is shown in Fig. 2. Corresponding to input PWM signals  $S_i \in \{0, 1\}$  in the voltage domain, each switched-current source (SCS) outputs current  $I_i$  when  $S_i = 1$ . An SCS can be replaced by a combination of a resistor and a diode if the nonlinearity in charging characteristics can be ignored. The weight is expressed by current  $I_i$ , which can be controlled by the gate voltage if an SCS consists of a field-effect transistor (FET). If the FET consists of a flash memory device or a ferroelectric-gate FET, current  $I_i$  can be changed and stored with nonvolatility, which means that the weights can be set arbitrarily and modified with an on-chip learning mechanism.

The total charge amount  $Q$  stored at the node of capacitor  $C$  charged by  $N$  SCSs with inputs  $S_i$ , each of which has a pulse width of  $W_i$ , is expressed by

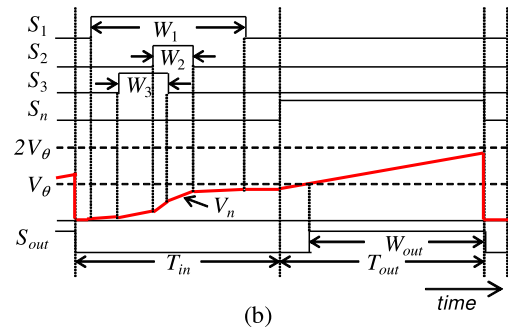
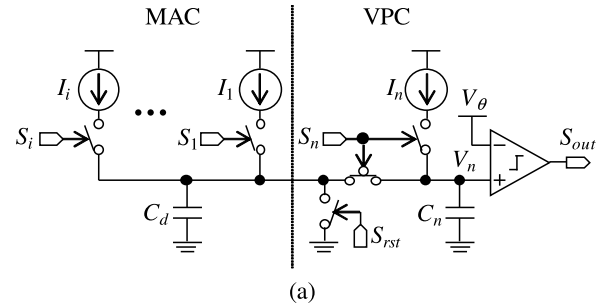
$$Q = \sum_{i=1}^N W_i I_i, \quad (1)$$

where  $Q$  can be considered as the weighted-sum calculation result with weight  $I_i$  and input  $W_i$ . The node voltage of  $C$ ,  $V_c$ , is given by  $V_c = Q/C$ , and the energy consumption  $E$  of this charging/discharging process is given by  $E = CV_c V_{dd}$ , where  $V_{dd}$  is a supply voltage of SCSs.

The weighted-sum calculation circuit and a timing diagram of its operation are shown in Fig. 3. Here, we consider this operation as a weighted-sum calculation with the same signed weighting. The circuit consists of a weighted-sum calculation or MAC part and a voltage-pulse conversion (VPC) part. The MAC part consists of SCSs corresponding with inputs, and is accompanied by the parasitic wiring capacitance  $C_d$ . The VPC part consists of an SCS, two switches, and a comparator with an input capacitance  $C_n$ . Since the parasitic capacitances  $C_d$  and  $C_n$  are inevitably included in the circuit, to minimize the energy consumption for the operation, the charged capacitance  $C$ , which is equal to  $C_d + C_n$ , should be as small as possible.

A possible circuit configuration based on the TACT-PWM approach is as follows. The PWM inputs are given in the input period  $T_{in}$ ;  $\forall i, W_i \leq T_{in}$ , which is arbitrarily defined. If the node voltage  $V_c$  at the timing of the end of this input period is denoted by  $V_{mac}$ ,

$$V_{mac} = \frac{Q}{C_d + C_n} = \frac{1}{C_d + C_n} \sum_{i=1}^N W_i I_i. \quad (2)$$



**FIGURE 3.** Weighted-sum calculation circuit model with the same signed weighting: (a) circuit diagram and (b) timing diagram.

In the VPC part, the output PWM signal  $S_{out}$  with pulse width  $W_{out}$  is generated during the output period  $T_{out}$ . In this operation, capacitance  $C$  is charged up by the SCS with current  $I_n$ . To minimize the energy consumption in this operation, the VPC part is separated from the MAC part by  $S_n$ , and only  $C_n$  can be charged up to the threshold voltage  $V_\theta$  of the comparator. In this case, to meet the condition that  $0 \leq W_{out} \leq T_{out}$ , the current  $I_n$  is given by

$$I_n = \frac{C_n V_\theta}{T_{out}}, \quad (3)$$

which means that the node voltage  $V_n$  increases with the slope of  $V_\theta/T_{out}$ . When  $V_n > V_\theta$ , the comparator output  $S_{out} = 1$ , and after the end of the output period  $V_n$  is reset by  $S_{rst}$  at the resting state, which is usually zero. Thus, the pulse width of the output signal as a result of weighted-sum calculation is given by

$$W_{out} = \frac{V_{mac}}{V_\theta} T_{out} \quad (4)$$

$$= \frac{T_{out}}{(C_d + C_n)V_\theta} \sum_{i=1}^N W_i I_i, \quad (5)$$

where it is assumed that  $0 \leq Q \leq (C_d + C_n)V_\theta$ .

The parameters  $T_{out}$  and  $V_\theta$  are determined with some constraints in the circuit design, and the maximum MAC value is determined by adjusting  $W_i$  and  $I_i$  so that Eqs. (2), (4), and (5) are satisfied. The parameters  $T_{in}$ ,  $T_{out}$  and  $V_\theta$  are fixed and determined by the required specifications (energy efficiency and calculation precision);  $T_{in}$  and  $T_{out}$  should be as short as possible to reduce the operating time of the comparator, which leads to a reduction of the energy consumption of the comparator and an improvement of the calculation performance. However, the minimum values of  $T_{in}$  and  $T_{out}$  depend

on the time resolution of the peripheral circuits that treat input and output signals. For example, if the time resolution is about 4 ns, and if the precision is required to be 6 bits,  $T_{in}$  and  $T_{out}$  should be more than 256 ns. The threshold voltage of the comparator,  $V_{\theta}$ , is limited by the supply voltage  $V_{dd}$ , typically  $V_{\theta} < V_{dd}/2$ , and reducing  $V_{dd}$  leads to a reduction in the energy consumption of the whole circuit. However, reducing  $V_{\theta}$  leads to a reduction in the signal-to-noise ratio of output signals, which means lowering the calculation precision.

If the same input line structures are used for both the positive and negative weights, the denominator of Eq. (5) is common. Thus, positive and negative weighted calculations are performed separately in the different lines, and by subtracting  $W_{out}$  for negative weighting from that for positive weighting, the total calculation result is obtained as follows:

$$W_{out}^+ - W_{out}^- = \frac{T_{out}}{(C_d + C_n)V_{\theta}} \left[ \sum_{i=1}^{N^+} W_i^+ I_i^+ - \sum_{i=1}^{N^-} W_i^- I_i^- \right], \quad (6)$$

$$N = N^+ + N^-, \quad (7)$$

where  $W_{out}^{\pm}$  are the pulse widths of output signals with positive and negative weighting, respectively. Since the obtained result can be fed into the next circuit corresponding to the next layer of the network via nonlinear transformation operation, the calculations for ANNs can be achieved.

The total energy consumption for the MAC calculation is expressed as follows:

$$E_{cal} = E_{mac} + E_{vpc}, \quad (8)$$

$$E_{mac} = C_d V_{mac} V_{dd} + \sum_{i=1}^N E_i, \quad (9)$$

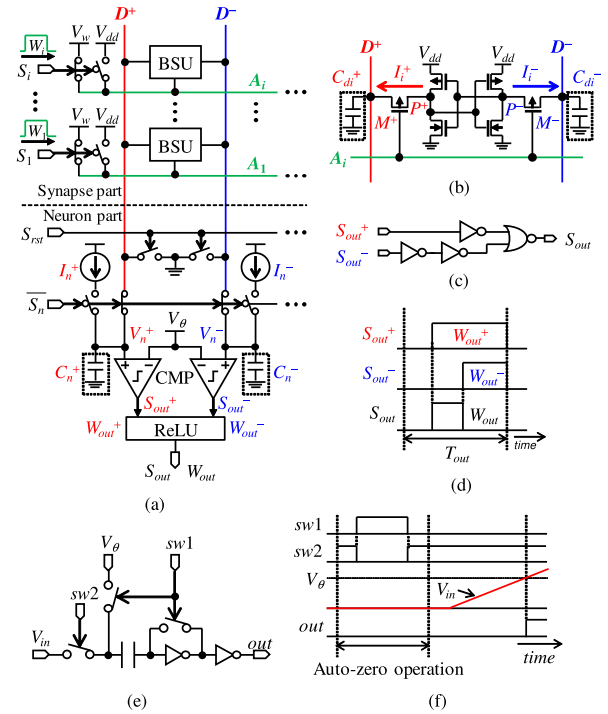
$$E_{vpc} = C_n (V_{mac} + V_{\theta}) V_{dd} + E_n + \int_0^{T_{in}+T_{out}} P_{cmp}(t) dt, \quad (10)$$

where  $E_{mac}$  and  $E_{vpc}$  are the energy consumptions of the MAC and VPC parts,  $E_i$  and  $E_n$  are those for the switching of the SCS at each MAC part  $i$  and for the switching of the SCS at the VPC part, respectively, and  $P_{cmp}(t)$  is the power consumption of the comparator.

### III. CMOS BinaryConnect NETWORK CIRCUIT BASED ON THE TACT-PWM APPROACH

On the basis of our TACT-PWM circuit approach, a CMOS circuit using an SRAM cell array structure was developed as shown in Fig. 4 (a). This circuit implements a BinaryConnect neural network, which uses analog input values while the weights are binary [22].

This circuit consists of a synapse part and a neuron part. The synapse part consists of an SRAM cell array, and each synapse circuit operates as two MAC circuits. Unlike the ordinary SRAM circuits proposed in the concept of in-memory computing, our SRAM cell circuit outputs very low current on the order of nano-amperes to guarantee the time constant in the TACT approach [36], [37], and therefore the p-type MOS



**FIGURE 4. BinaryConnect neural network circuit based on the TACT-PWM approach: (a) schematic diagram, (b) binary synapse unit (BSU) circuit, (c) ReLU function circuit, (d) timing diagram of the ReLU function circuit, (e) comparator (CMP) circuit, and (f) timing diagram of the comparator.**

field effect transistors (pMOSFETs)  $M^{\pm}$  supply subthreshold currents to the *dendrite* lines  $D^{\pm}$  based on the input from the *axon* lines  $A_i$ , where *axon* and *dendrite* are neuroscientific terms referencing biological neurons. As shown in Fig. 4 (a), when an input PWM pulse  $S_i$  with a pulse width  $W_i$  is fed into the circuit, the voltage of the line  $A_i$  is changed from  $V_{dd}$  to  $V_w$ , and pMOSFETs  $M^{\pm}$  are turned on during the time span of  $W_i$ .

In the neuron part, two VPC circuits perform the positive and negative weighting calculations, respectively, and the subtraction result is obtained by a rectified-linear-unit (ReLU) function circuit. A detailed explanation follows.

#### A. SYNAPSE PART

In the synapse part, each SRAM cell shown in Fig. 4 (b), hereinafter called a binary synapse unit (BSU), performs binary weighting, when receiving an input pulse  $S_i$  as the gate voltage of the pMOSFET  $M^{\pm}$  to make it operate in the subthreshold region. To perform this operation, it is necessary that the SRAM cell be set at a 0 or 1 state based on the training result in a BinaryConnect network.

The BSU has three functions: one-bit memory, a switched current source, and a selector. The one-bit memory function is achieved at the flip-flop, which stores the binary weight  $w_i \in \{+1, -1\}$  by setting voltages  $V_p^+$  and  $V_p^-$ , as follows:

$$w_i = \begin{cases} +1 & \text{if } (V_p^+, V_p^-) = (V_{dd}, 0) \\ -1 & \text{if } (V_p^+, V_p^-) = (0, V_{dd}) \end{cases}. \quad (11)$$

The switched current source with a selector is realized by the pMOSFETs  $M^{\pm}$  that are connected to the dendrite lines

$D^\pm$ , respectively. Since the pMOSFETs  $M^\pm$  operate in the subthreshold region, their drain currents  $I_i^\pm$  are expressed as follows:

$$I_i^\pm \approx I_0 \exp(V_P^\pm - V_{Ai}) \quad (12)$$

$$V_{Ai} = \begin{cases} V_{dd} & \text{if } S_i = 0 \\ V_w & \text{if } S_i = 1 \end{cases}, \quad (13)$$

where  $I_0$  is a constant,  $V_{Ai}$  is the voltage of axon line  $A_i$ , and  $V_w$  is the constant gate voltage for subthreshold operation. For example, if synapse  $i$  has positive weight ( $w_i = 1$ ) and  $S_i = 1$ , then  $(V_P^+, V_P^-) = (V_{dd}, 0)$ , and  $I_w^+ \approx I_0 \exp(V_{dd} - V_w)$ , and  $I_w^- \approx 0$ .

The currents  $I_i^\pm$  flowing from each BSU to the lines  $D^\pm$  used for MAC calculations consist of two parts: a subthreshold current of  $M^\pm$  and leakage currents, which are the junction leakage and gate-induced drain leakage (GIDL). However, the subthreshold currents can be significantly higher than these leakage currents. The leakage currents can be considered as offsets and somewhat canceled out in the subtraction operation between  $V^\pm$  for executing the ReLU function.

### B. NEURON PART

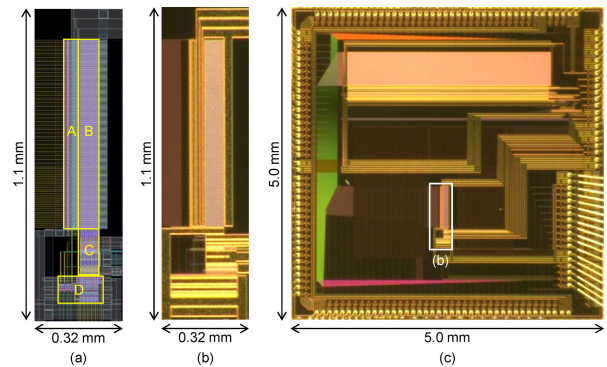
In the neuron circuit, dendrite lines are initialized and reset at ground level by  $S_{rst}$  before inputting signals  $S_i$  to the synapse part. Next, input PWM signals are given during input time period  $T_{in}$ , and the capacitances  $C_{di}$  and  $C_n$  are charged. Then, the dendrite lines are separated from the neuron parts with  $S_n$ . At the same time, the current source  $I_n$  is connected to the capacitance  $C_n$ , and thus  $C_n$  is charged. When the node voltage of  $C_n$ ,  $V_n^\pm$ , reaches the threshold voltage of the comparator, the output signal  $S_{out}^\pm$  is generated. A set of output signals  $S_{out}^\pm$  are fed into the ReLU function circuit, which simply consists of logic circuits, as shown in Fig. 4(c), and the output PWM signal is only generated when  $W_{out}^+ > W_{out}^-$ , as shown in Fig. 4(d).

The comparator circuit CMP and its timing diagram are shown in Figs. 4(e) and 4(f), respectively. In this design, a clocked CMOS inverter was used as a CMP, where the  $V_\theta$  setting and compensation of the MOSFET threshold voltage variation are achieved with charges stored in a capacitor shown in Fig. 4(e) by the *auto-zero* operation shown in Fig. 4(f).

### IV. VLSI CHIP DESIGN AND MEASUREMENT RESULTS

Using TSMC 250 nm CMOS technology, we designed and fabricated a CMOS VLSI chip of a single-layer neural network circuit with ten neurons each of which has 100 synapses based on the TACT-PWM approach. The layout results and microphotographs are shown in Fig. 5.

The whole chip photograph is shown in Fig. 5(c), where other unrelated circuits with long wires are also shown. The proposed circuit is shown inside the white rectangle (Fig. 5(b)). When multi-layer networks are constructed, the output of the ReLU circuit is connected to the input of the next layer circuits, which can be located nearby. However, if the wiring distance becomes long, the wiring capacitance



**FIGURE 5.** VLSI layout results of a  $100 \times 10$  BinaryConnect neural network: (a) layout result, (b) microphotograph of the circuit, and (c) chip microphotograph. A: Switch and buffer array for axon lines; B: BSU array; C: neuron array; and D: buffer array for dendrite lines.

**TABLE 1.** Measurement conditions and results of the fabricated VLSI chip.

Number of synapses	$100 \times 10$
Operations per synapse	2 (MAC)
Number of neurons	10
Input pulse width	300 ns - 2 $\mu$ s
Output pulse width	300 ns - 2 $\mu$ s
Supply voltage $V_{dd}$	1 V
Threshold voltage $V_\theta$	0.2 - 0.4 V
Operation freq.	120 kHz - 290 kHz
Throughput	0.24 GOPS - 0.59 GOPS
Power consumption	1.6 $\mu$ W - 1.9 $\mu$ W
Energy efficiency	150 - 300 TOPS/W

increases, which leads to an energy consumption increase. Nonetheless, the delay in signal transmission causes no precision degradation because the delays at the rising and trailing edges of the PWM signals can be considered equal.

### A. EVALUATION OF ENERGY EFFICIENCY AND CALCULATION PRECISION

We evaluated the fabricated chip under various conditions, and calculated the energy efficiency based on the measurement results, as shown in Table 1. Fig. 6 shows the energy consumptions per operation measured under typical conditions. Here, leakage currents in the SRAM cells are also included in the power consumption. The highest energy efficiency obtained from the measurement was 300 TOPS/W, as shown in Fig. 6C.

The proposed circuit requires three types of bias voltages:  $V_w$ ,  $V_\theta$ , and a bias voltage for setting  $I_n$ , which have to be supplied from external voltage sources, and the power consumption of these voltage sources are not included in Table 1. However, these bias voltages can be shared by all the same component circuits. We estimated the power consumption when commercial integrated circuits (ICs) were used. For example, in the evaluation of the test chip, we used three low-noise power supply ICs (LT3042), and the total power consumption was about 11 mW. Thus, when a very large number of MAC circuits (more than 6 million) are used, the power consumption of external bias voltage supplies can be negligible.

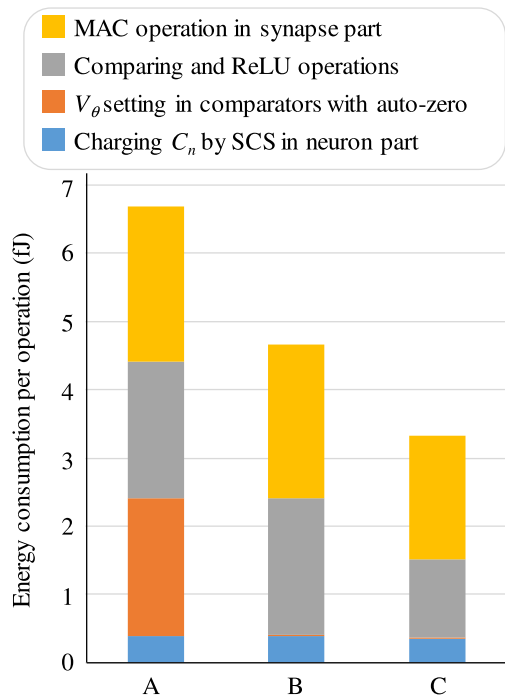
The comparison results among some typical AI processors are shown in Table 2. In AI processors targeted mainly

**TABLE 2. Comparison of state-of-the-art AI processors.**

Reference	[26]	[25]	[17]	[18]	[18] (N)	This work	This work (N)
Technology node (nm)	65	65	65	7	(65)	250	(65)
$p_{65nm}$	-	-	-	-	200/40	-	200/800
Precision of Weight/Input	1b/1b	1b/6b	8b/8b	4b/4b	4b/4b	1b/analog( $\sim$ 6b)	1b/analog( $\sim$ 6b)
Throughput/area (GOPS/mm <sup>2</sup> )	1498	(33)* <sup>1</sup>	(10.6)* <sup>1</sup>	116400	(4656)	5.9	(94)
Max. Energy Efficiency (TOPS/W)	866	51.3	6.25	351	(70)	300	(1200)

(N): estimation results with conversion into 65 nm technology node, and  $p_{65nm}$  is a shrinkage ratio to the 65 nm node based on line pitches in metal layers.

\*1: core circuit footprint was estimated from a die microphotograph.



**FIGURE 6.** A:  $V_\theta = 0.4$  V,  $\forall i$   $W_i = 2$   $\mu$ s, auto-zeroing frequency per weighted sum operation:  $f_{az} = 1$ , energy efficiency: 150 TOPS/W, jitter:  $3\sigma/W_{out\_max} = 0.010$ ; B:  $V_\theta = 0.4$  V,  $\forall i$   $W_i = 2$   $\mu$ s,  $f_{az} = 1/100$ , energy efficiency: 210 TOPS/W, jitter:  $3\sigma/W_{out\_max} = 0.011$ ; C:  $V_\theta = 0.2$  V,  $\forall i$   $W_i = 0.3$   $\mu$ s,  $f_{az} = 1/100$ , energy efficiency: 300 TOPS/W, jitter:  $3\sigma/W_{out\_max} = 0.021$ .

for edge computing, the most important performance measure is energy efficiency, which is independent of throughput. The energy efficiency of 300 TOPS/W in this work is six times as high as that of a state-of-the-art in-memory-computing AI processor implementing the same BinaryConnect model, while the VLSI technology node used in this work is four-generation older than that in the AI processor [25], as shown in Table 2. One of the latest AI processors fabricated using a 7 nm-node VLSI technology [18] is also included in Table 2. It is noted that, even compared with this processor, the energy efficiency of our present processor is comparable.

Furthermore, in Table 2, we roughly estimated the performance of this work virtually at a 65 nm node. It is rather difficult to compare AI processors with different circuit approaches and using different VLSI technologies, but we added in Table 2 estimation results with conversion into 65 nm technology, where  $p_{65nm}$  is the shrinkage ratio

estimated with line pitches in interconnection metal layers. The energy efficiency of our approach is mainly determined by the operation voltage and capacitance, such as  $V_{mac}$ ,  $C_d$  and  $C_n$ , as shown in Eqs. (9) and (10). The operation voltage is around 1 V, which is almost unchanged even in technology nodes of 65 nm and 7 nm, although the voltage lowers gradually. Therefore, the energy efficiency is determined by the capacitance, which is mainly given by the parasitic capacitance between interconnection wires in the metal layers. Therefore, the shrinkage ratio  $p_{65nm}$  is given by the second metal line pitch in each technology node [42]. On the other hand, the throughput is determined by the circuit footprint density per unit chip area, which is given by  $(1/p_{65nm})^2$ .

Using this ratio with the scaling trend down to the 7 nm node, the throughput is converted proportional to  $(1/p_{65nm})^2$ , and the energy efficiency is converted proportional to  $1/p_{65nm}$ . As shown in Table 2, if we fabricate a TACT-PWM-based AI processor using the same VLSI technology as in the digital AI processors, we will obtain an energy efficiency of more than 1,000 TOPS/W or 1 POPS/W (Peta-OPS/W).

It is noted that the low throughput, 0.24-0.59 GOPS, as shown in Table 1, is due to the mature VLSI technology, such as TSMC 250 nm CMOS technology, used in this implementation and the small number of MAC circuits integrated in a chip. The throughput can be improved by increasing the number of processing elements in a chip, which can easily be achieved by using more advanced fabrication technology or increasing the chip area. Regarding the throughput per area converted to the same technology, this work is superior to another AI processor implementing the BinaryConnect model, as shown in Table 2.

Measurement results of the input-output relationship in weighted-sum calculations operations at one neuron with 100 synapses are shown in Fig. 7. As shown in Fig. 7(a), a weighted-sum operation was approximately achieved, and sufficient linearity was obtained. From Fig. 7(b), the deviations in the time domain are  $\pm 20$  ns, and this means that the precision of the calculation is about  $\pm 1$  % because of the maximum pulse width being 2  $\mu$ s.

Offsets and scattering of weighting are clearly observed in Fig. 7(a). These nonidealities are mainly caused by the following variations: variations in parasitic capacitance  $C^\pm$  and variations in the threshold voltages of MOSFETs  $M^\pm$  operating in the subthreshold region in BSUs. Variations in  $C_n^\pm$  and  $I_n^\pm$  in the neuron part also create errors in the pulse generation operation. If the proposed approach is implemented using

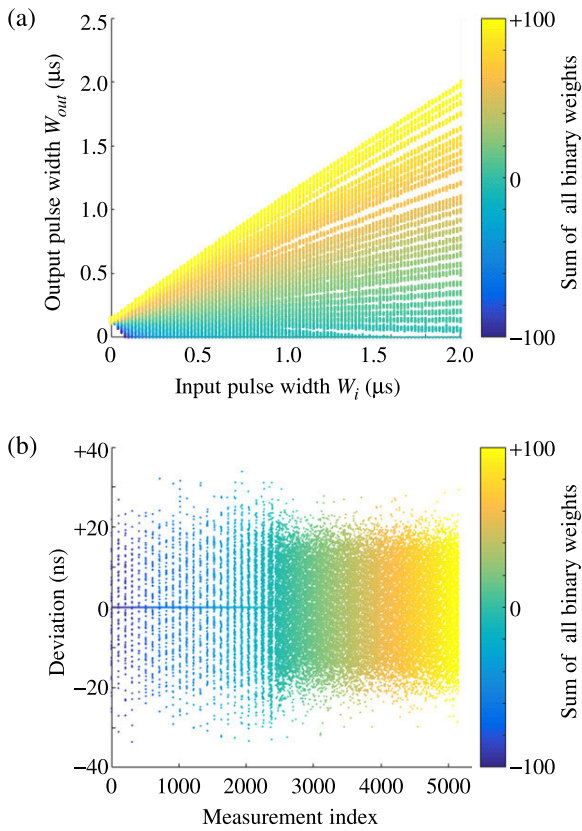


FIGURE 7. Measurement results of input-output characteristics: (a) averaged output pulse width and (b) deviation.

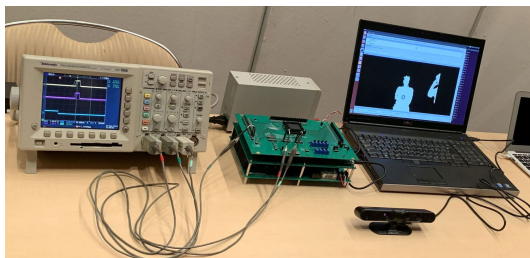


FIGURE 8. Human detection system using the fabricated chip.

a more advanced VLSI technology than the 250 nm node used in our current implementation, these errors will increase, because the variations of threshold voltages of MOSFETs will increase with an advanced scaling. It is known that such variations are in proportion to  $t_{ox}/\sqrt{LW}$ , where  $t_{ox}$  is the thickness of the gate oxide film, and  $L$  and  $W$  are the length and width of the gate electrode of a MOSFET, respectively [43], [44]. If we design the processor with a simple scaling from 250 nm to 65 nm node, which means a four-generation advance, the variations will only be two or three times larger by virtue of the different techniques used to suppress the increase of variations. Although these errors may lead to large current variations in the subthreshold operation of MOSFETs, these can be controlled by the values of  $L$  and  $W$ , and therefore, they are not fatal for some recognition tasks. Furthermore, if analog memory devices



FIGURE 9. Home service robot having the system including the fabricated chip.

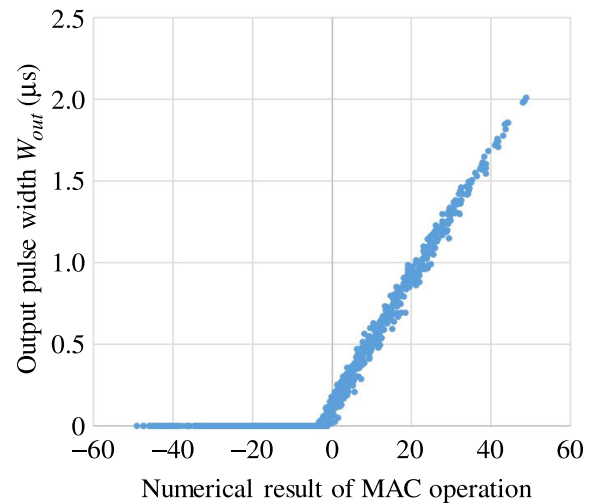
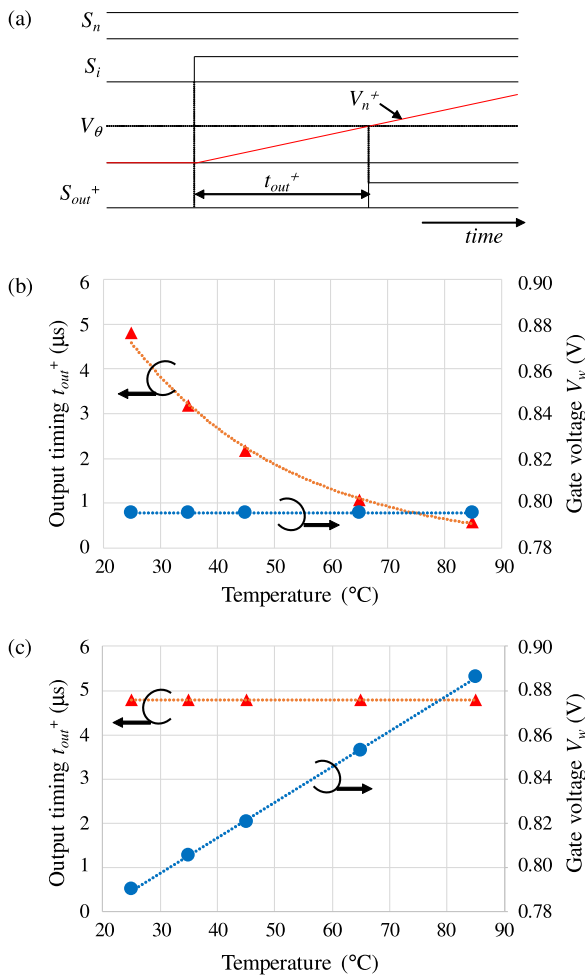


FIGURE 10. Measurement results of output pulse widths for the combination of random weights and inputs. Timing jitters were decreased by averaging the output signals for 50 measurement results. The horizontal axis shows the numerical calculation values of  $\sum_{i=1}^{N=50} w_i \cdot W_i/T_i$ , where  $w_i \in \{+1, -1\}$  and  $0 \leq W_i/T_{in} \leq 1$ .

such as ferroelectric-gate FETs are used as  $M^\pm$  in BSUs and MOSFETs operating for the current sources  $I_n^\pm$ , such variations can be compensated for by adjusting the threshold voltages, and the recognition success rate will be improved.

We applied the fabricated chip controlled by a microcomputer to a home service robot, and evaluated its classification performance in the human recognition task. The system was proposed and demonstrated in a live demonstration at an international conference, as shown in Fig. 8 [45]. Fig. 9 shows a robot called HSR [46] that has the human recognition system including the fabricated chip. The fabricated chip was



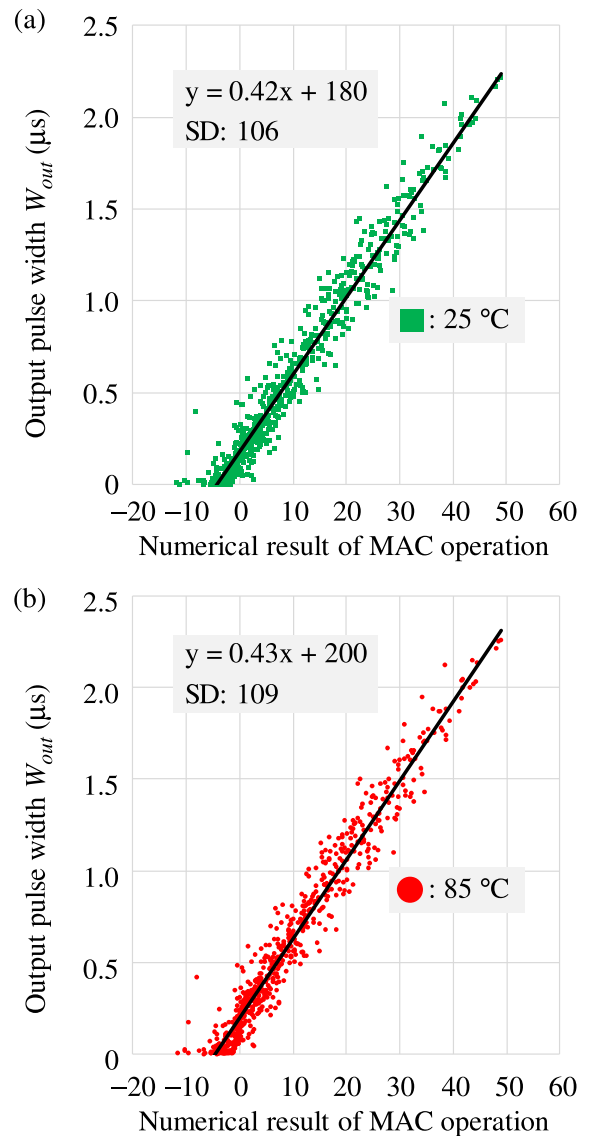
**FIGURE 11.** Measurement results of the temperature characteristics of the TACT-PWM circuit: (a) measurement condition, (b) temperature dependence of the output with constant  $V_w$ , and (c) temperature dependence with changing  $V_w$  in order to make  $t_{out}^+$  constant.

used as weighted-sum operation units at the last stage of a classifier. The detection success rate of the classifier was 86 % of that obtained by numerical simulation.

The measurement results of the output pulse width as a function of the weighted-sum calculation results followed by the ReLU function in one neuron with 100 synapses are shown in Fig. 10. The average error was 1.5 %, and the maximum error was about 8 %. This error will be decreased by adjusting the deviations of the threshold voltages of MOSFETs operating in the subthreshold region by using analog memory devices.

**B. TEMPERATURE DEPENDENCE OF THE TACT-PWM CIRCUIT**

The temperature characteristics of the proposed calculation circuits are very important, because the ambient temperature crucially affects the analog circuit operation and calculation precision. Therefore, we measured the temperature dependence of the fabricated TACT-PWM circuit with 100 inputs and one output.



**FIGURE 12.** Measurement results of output pulse widths with random weights and inputs, where the conditions are the same as in Fig. 10 except temperature,  $V_w$ , and  $V_{in}$ : (a)  $T = 298\text{ K}$ ,  $V_w = 0.79\text{ V}$ , and  $V_{in} = 0.44\text{ V}$ ; (b)  $T = 358\text{ K}$ ,  $V_w = 0.86\text{ V}$ , and  $V_{in} = 0.49\text{ V}$ .

Fig. 11(a) shows the measurement conditions of the temperature characteristics of the TACT-PWM circuit. Here, the switched current source of the neuron part was always ‘OFF’ ( $S_n = 1$ ), and therefore the output of the neuron was inverted only by currents from the synapse part. All weights of the synapse part were set at positive ( $w_i = +1$ ), and all inputs were set at ‘1’ ( $S_i = 1$ ), and we observed an inversion timing  $t_{out}^+$  of the positive output  $S_{out}^+$ . It is noted that we used a clocked CMOS inverter as the comparator in the neuron part, in which the threshold voltage deviation by temperature change as well as device mismatch are compensated for by the auto-zero operation. Therefore, we were able to measure only the temperature characteristics of the synapse part of the TACT-PWM circuit by observing  $t_{out}^+$  with different temperatures.



The temperature characteristics of the synapse part are shown in Fig. 11(b) and (c), where the temperature is changed from 25°C to 85°C; Fig. 11(b) shows the measurement results with a constant  $V_w$ , and Fig. 11(c) shows those with changing  $V_w$  in order to make  $t_{out}^+$  constant. These results show that the temperature characteristics of  $I_i^+$  can be compensated for by changing  $V_w$  linearly.

Fig. 12 shows the measurement results of output pulse widths with random weights and inputs at temperatures of 25°C (Fig. 12(a)) and 85°C (Fig. 12(b)). Here,  $V_w$  was set based on the results shown in Fig. 11(c), and  $I_n^\pm$  was set at a constant value by adjusting  $V_{in}$  in Fig. 4(e). The approximately straight lines of Fig. 12(a) and (b) are almost equal. Therefore, these results show that effective temperature compensation is achieved by adjusting  $V_w$  based on Fig. 11(c).

## V. CONCLUSION

In this paper, we proposed a time-domain weighted-sum calculation model based on the TACT-PWM approach with an activation function of ReLU. We also proposed VLSI circuits based on the TACT approach to implement a BinaryConnect model with extremely low energy consumption. A high energy efficiency of 300 TOPS/W was achieved by the fabricated CMOS VLSI circuit with binary weights using 250-nm CMOS VLSI technology. If we use a more advanced VLSI fabrication technology, which achieves lower parasitic capacitance, the energy efficiency will be further improved to over 1,000 TOPS/W.

The fabricated circuit had limited calculation precision, which was mainly due to the characteristic variations of subthreshold operation in MOSFETs. To improve the calculation precision and compensate for such variations, it is necessary to introduce analog memory devices. We also evaluated the temperature characteristics of the circuit by measuring the fabricated chip, and suggested that the effects by a temperature change can be compensated for by adjusting the bias voltage.

As for the neuron part, the measurement results of the fabricated VLSI chip suggest that the energy consumption of this part is comparable to that of the whole synapse part with 100 inputs. Therefore, it is also necessary to refine a comparator circuit with much lower power consumption to improve the energy efficiency of the whole calculation circuit.

## ACKNOWLEDGMENT

The authors thank Kazumasa Yanagisawa for their invaluable discussion about VLSI technology scaling and device variations. Part of the work was carried out under a project, JPNP16007, commissioned by the New Energy and Industrial Technology Development Organization (NEDO), and the Collaborative Research Project of the Institute of Fluid Science, Tohoku University. The circuit design was supported by VLSI Design and Education Center (VDEC), the University of Tokyo in collaboration with Cadence Design Systems, Inc., Mentor Graphics, Inc., and Synopsys, Inc.

## REFERENCES

- [1] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proc. IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.
- [2] D. C. Cireşan, U. Meier, L. M. Gambardella, and J. Schmidhuber, "Deep, big, simple neural nets for handwritten digit recognition," *Neural Comput.*, vol. 22, no. 12, pp. 3207–3220, Dec. 2010.
- [3] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2012, pp. 1097–1105.
- [4] C. Farabet, C. Couprie, L. Najman, and Y. LeCun, "Learning hierarchical features for scene labeling," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 8, pp. 1915–1929, Aug. 2013.
- [5] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 1–9.
- [6] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.
- [7] J. Sim, J.-S. Park, M. Kim, D. Bae, Y. Choi, and L.-S. Kim, "A 1.42TOPS/W deep convolutional neural network recognition processor for intelligent IoT systems," in *IEEE Int. Solid-State Circuits Conf. (ISSCC) Dig. Tech. Papers*, Jan./Feb. 2016, pp. 264–265.
- [8] B. Moons, R. Uytterhoeven, W. Dehaene, and M. Verhelst, "Envision: A 0.26-to-10TOPS/W subword-parallel dynamic-voltage-accuracy-frequency-scalable convolutional neural network processor in 28nm FDSOI," in *IEEE Int. Solid-State Circuits Conf. (ISSCC) Dig. Tech. Papers*, Feb. 2017, pp. 246–247.
- [9] D. Shin, J. Lee, J. Lee, and H. Yoo, "DNPU: An 8.1TOPS/W reconfigurable CNN-RNN processor for general-purpose deep neural networks," in *IEEE Int. Solid-State Circuits Conf. (ISSCC) Dig. Tech. Papers*, Feb. 2017, pp. 240–241.
- [10] W.-S. Khwa, J.-J. Chen, J.-F. Li, X. Si, E.-Y. Yang, X. Sun, R. Liu, P.-Y. Chen, Q. Li, S. Yu, and M.-F. Chang, "A 65nm 4Kb algorithm-dependent computing-in-memory SRAM unit-macro with 2.3ns and 55.8TOPS/W fully parallel product-sum operation for binary DNN edge processors," in *IEEE Int. Solid-State Circuits Conf. (ISSCC) Dig. Tech. Papers*, Feb. 2018, pp. 496–498.
- [11] L. Fick, D. Blaauw, D. Sylvester, S. Skrzyniarz, M. Parikh, and D. Fick, "Analog in-memory subthreshold deep neural network accelerator," in *Proc. IEEE Custom Integr. Circuits Conf. (CICC)*, Apr. 2017, pp. 1–4.
- [12] E. H. Lee and S. S. Wong, "A 2.5GHz 7.7TOPS/W switched-capacitor matrix multiplier with co-designed local memory in 40nm," in *IEEE Int. Solid-State Circuits Conf. (ISSCC) Dig. Tech. Papers*, Jan./Feb. 2016, pp. 418–419.
- [13] D. Miyashita, S. Kousai, T. Suzuki, and J. Deguchi, "A neuromorphic chip optimized for deep learning and CMOS technology with time-domain analog and digital mixed-signal processing," *IEEE J. Solid-State Circuits*, vol. 52, no. 10, pp. 2679–2689, Oct. 2017.
- [14] M. R. Mahmoodi and D. Strukov, "An ultra-low energy internally analog, externally digital vector-matrix multiplier based on NOR flash memory technology," in *Proc. 55th ACM/ESDA/IEEE Design Autom. Conf. (DAC)*, Jun. 2018, pp. 1–6.
- [15] D. Milojicic, K. Bresniker, G. Campbell, P. Faraboschi, J. P. Strachan, and S. Williams, "Computing in-memory, revisited," in *Proc. IEEE 38th Int. Conf. Distrib. Comput. Syst. (ICDCS)*, Jul. 2018, pp. 1300–1309.
- [16] N. Verma, H. Jia, H. Valavi, Y. Tang, M. Ozatay, L. Chen, B. Zhang, and P. Deaville, "In-memory computing: Advances and prospects," *IEEE Solid-State Circuits Mag.*, vol. 11, no. 3, pp. 43–55, Summer 2019.
- [17] S. K. Gonugondla, M. Kang, and N. Shanbhag, "A 42pj/decision 3.12TOPS/W robust in-memory machine learning classifier with on-chip training," in *IEEE Int. Solid-State Circuits Conf. (ISSCC) Dig. Tech. Papers*, Feb. 2018, pp. 490–492.
- [18] Q. Dong, M. E. Sinangil, B. Erbagci, D. Sun, W.-S. Khwa, H.-J. Liao, Y. Wang, and J. and Chang, "A 351TOPS/W and 372.4GOPS compute-in-memory SRAM macro in 7nm FinFET CMOS for machine-learning applications," in *IEEE Int. Solid-State Circuits Conf. (ISSCC) Dig. Tech. Papers*, Feb. 2020, pp. 242–243.
- [19] T. Morie and Y. Amemiya, "An all-analog expandable neural network LSI with on-chip backpropagation learning," *IEEE J. Solid-State Circuits*, vol. 29, no. 9, pp. 1086–1093, Sep. 1994.

- [20] G. Indiveri, "Computation in neuromorphic analog VLSI systems," in *Proc. Italian Workshop Neural Nets (WIRN)*, 2001, pp. 3–20.
- [21] X. Guo, F. M. Bayat, M. Prezioso, Y. Chen, B. Nguyen, N. Do, and D. B. Strukov, "Temperature-insensitive analog vector-by-matrix multiplier based on 55 nm NOR flash memory cells," in *Proc. IEEE Custom Integr. Circuits Conf. (CICC)*, Apr. 2017, pp. 1–4.
- [22] M. Courbariaux, Y. Bengio, and J.-P. David, "BinaryConnect: Training deep neural networks with binary weights during propagations," in *Proc. Adv. Neural Inf. Process. Syst.*, 2015, pp. 3123–3131.
- [23] I. Hubara, M. Courbariaux, D. Soudry, R. El-Yaniv, and Y. Bengio, "Quantized neural networks: Training neural networks with low precision weights and activations," *J. Mach. Learn. Res.*, vol. 18, no. 1, pp. 6869–6898, 2017.
- [24] J. Lee, C. Kim, S. Kang, D. Shin, S. Kim, and H.-J. Yoo, "UNPU: A 50.6TOPS/W unified deep neural network accelerator with 1b-to-16b fully-variable weight bit-precision," in *IEEE Int. Solid-State Circuits Conf. (ISSCC) Dig. Tech. Papers*, Feb. 2018, pp. 218–220.
- [25] A. Biswas and A. P. Chandrakasan, "CONV-SRAM: An energy-efficient SRAM with in-memory dot-product computation for low-power convolutional neural networks," *IEEE J. Solid-State Circuits*, vol. 54, no. 1, pp. 217–230, Jan. 2019.
- [26] H. Valavi, P. J. Ramadge, E. Nestler, and N. Verma, "A 64-tile 2.4-Mb In-Memory-Computing CNN accelerator employing charge-domain compute," *IEEE J. Solid-State Circuits*, vol. 54, no. 6, pp. 1789–1799, Jun. 2019.
- [27] W. Maass, "Fast sigmoidal networks via spiking neurons," *Neural Comput.*, vol. 9, no. 2, pp. 279–304, Feb. 1997.
- [28] W. Maass, "Computing with spiking neurons," in *Pulsed Neural Networks*, W. Maass and C. M. Bishop, Eds. Cambridge, MA, USA: MIT Press, 1999, pp. 55–85.
- [29] T. Morie, Y. Sun, H. Liang, M. Igarashi, C.-H. Huang, and S. Samukawa, "A 2-dimensional Si nanodisk array structure for spiking neuron models," in *Proc. IEEE Int. Symp. Circuits Syst. (ISCAS)*, May 2010, pp. 781–784.
- [30] T. Tohara, H. Liang, H. Tanaka, M. Igarashi, S. Samukawa, K. Endo, Y. Takahashi, and T. Morie, "Silicon nanodisk array with a fin field-effect transistor for time-domain weighted sum calculation toward massively parallel spiking neural networks," *Appl. Phys. Exp.*, vol. 9, no. 3, 2016, Art. no. 034201.
- [31] T. Morie, H. Liang, T. Tohara, H. Tanaka, M. Igarashi, S. Samukawa, K. Endo, and Y. Takahashi, "Spike-based time-domain weighted-sum calculation using nanodevices for low power operation," in *Proc. IEEE 16th Int. Conf. Nanotechnol. (IEEE-NANO)*, Aug. 2016, pp. 390–392.
- [32] M. Liu, L. R. Everson, and C. H. Kim, "A scalable time-based integrate-and-fire neuromorphic core with brain-inspired leak and local lateral inhibition capabilities," in *Proc. IEEE Custom Integr. Circuits Conf. (CICC)*, Apr. 2017, pp. 1–4.
- [33] N. Cao, M. Chang, and A. Raychowdhury, "A 65nm 1.1-to-9.1TOPS/W hybrid-digital-mixed-signal computing platform for accelerating model-based and model-free swarm robotics," in *IEEE Int. Solid-State Circuits Conf. (ISSCC) Dig. Tech. Papers*, Feb. 2019, pp. 222–224.
- [34] Z. Chen and J. Gu, "A scalable pipelined time-domain DTW engine for time-series classification using multibit time flip-flops with 140Giga-cell-updates/s throughput," in *IEEE Int. Solid-State Circuits Conf. (ISSCC) Dig. Tech. Papers*, Feb. 2019, pp. 324–326.
- [35] Z. Chen and J. Gu, "A time-domain computing accelerated image recognition processor with efficient time encoding and non-linear logic operation," *IEEE J. Solid-State Circuits*, vol. 54, no. 11, pp. 3226–3237, Nov. 2019.
- [36] Q. Wang, H. Tamukoh, and T. Morie, "Time-domain weighted-sum calculation for ultimately low power VLSI neural networks," in *Proc. Int. Conf. Neural Inf. Process. (ICONIP)*, 2016, pp. 240–247.
- [37] Q. Wang, H. Tamukoh, and T. Morie, "A time-domain analog weighted-sum calculation model for extremely low power VLSI implementation of multi-layer neural networks," 2018, *arXiv:1810.06819*. [Online]. Available: <http://arxiv.org/abs/1810.06819>
- [38] M. Bavandpour, M. Reza Mahmoodi, and D. B. Strukov, "Energy-efficient time-domain Vector-by-Matrix multiplier for neurocomputing and beyond," 2017, *arXiv:1711.10673*. [Online]. Available: <http://arxiv.org/abs/1711.10673>
- [39] M. Nagata, J. Funakoshi, and A. Iwata, "A PWM signal processing core circuit based on a switched current integration technique," *IEEE J. Solid-State Circuits*, vol. 33, no. 1, pp. 53–60, Jan. 1998.
- [40] M. Yamaguchi, G. Iwamoto, H. Tamukoh, and T. Morie, "An energy-efficient time-domain analog VLSI neural network processor based on a pulse-width modulation approach," 2019, *arXiv:1902.07707*. [Online]. Available: <http://arxiv.org/abs/1902.07707>
- [41] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," 2015, *arXiv:1512.03385*. [Online]. Available: <http://arxiv.org/abs/1512.03385>
- [42] H. Goto, *Process Node Comparison (Gate pitch/CPP x Minimum Metal Pitch + Fin Pitch)*. Accessed: Oct. 1, 2020. [Online]. Available: <https://pc.watch.impress.co.jp/video/pcw/docs/1187/086/p11.pdf>
- [43] M. J. M. Pelgrom, A. C. J. Duinmaijer, and A. P. G. Welbers, "Matching properties of MOS transistors," *IEEE J. Solid-State Circuits*, vol. 24, no. 5, pp. 1433–1440, Oct. 1989.
- [44] Y. Taur and T. H. Ning, *Fundamentals of Modern VLSI Devices*. New York, NY, USA: Cambridge Univ. Press, 1998.
- [45] M. Yamaguchi, G. Iwamoto, Y. Abe, Y. Tanaka, Y. Ishida, H. Tamukoh, and T. Morie, "Live demonstration: A VLSI implementation of time-domain analog weighted-sum calculation model for intelligent processing on robots," in *Proc. IEEE Int. Symp. Circuits Syst. (ISCAS)*, May 2019.
- [46] T. Yamamoto, K. Terada, A. Ochiai, F. Saito, Y. Asahara, and K. Murase, "Development of Human Support Robot as the research platform of a domestic mobile manipulator," *ROBOMECH J.*, vol. 6, Apr. 2019, Art. no. 4.



**MASATOSHI YAMAGUCHI** received the B.E. and M.E. degrees from Kyushu Institute of Technology, Japan, in 2015 and 2017, respectively. His main research interests include VLSI implementation of neural networks.



**GOKI IWAMOTO** received the B.E. and M.E. degrees from the Kyushu Institute of Technology, Japan, in 2017 and 2019, respectively. His research interests include VLSI implementation of neural networks.



**YUTA NISHIMURA** received the B.E. degree from the Kyushu Institute of Technology, Japan, in 2019, where he is currently pursuing the M.E. degree. His research interest includes VLSI implementation of neural networks.



**HAKARU TAMUKOH** (Member, IEEE) received the B.E. degree from Miyazaki University, Japan, in 2001, and the M.E. and Ph.D. degrees from the Kyushu Institute of Technology, Japan, in 2003 and 2006, respectively. He was a Postdoctoral Research Fellow of the 21st Century Center of Excellence Program at the Kyushu Institute of Technology, from 2006 to 2007. He was an Assistant Professor with the Tokyo University of Agriculture and Technology, from 2007 to 2013.

He is currently an Associate Professor with the Graduate School of Life Science and System Engineering, Kyushu Institute of Technology, Japan. His research interests include hardware/software complex systems, digital hardware design, neural networks, soft-computing, intelligent robotics, and signal processing.



**TAKASHI MORIE** (Member, IEEE) received the B.S. and M.S. degrees in physics from Osaka University, Osaka, Japan, in 1979 and 1981, respectively, and the Dr.Eng. degree from Hokkaido University, Sapporo, Japan, in 1996. From 1981 to 1997, he was a member of the Research Staff at Nippon Telegraph and Telephone Corporation (NTT). From 1997 to 2002, he was an Associate Professor with the Department of Electrical Engineering, Hiroshima University, Higashihiroshima, Japan. Since 2002, he has been a Professor with the Graduate School of Life Science and Systems Engineering, Kyushu Institute of Technology, Kitakyushu, Japan. His research interests include VLSI implementation of neural networks and new functional nanodevices.

• • •