



Back-Propagation Operation for Analog Neural Network Hardware with Synapse Components Having Hysteresis Characteristics

Michihito Ueda*, Yu Nishitani, Yukihiro Kaneko, Atsushi Omote

Advanced Research Division, Panasonic Corporation, Soraku, Kyoto, Japan

Abstract

To realize an analog artificial neural network hardware, the circuit element for synapse function is important because the number of synapse elements is much larger than that of neuron elements. One of the candidates for this synapse element is a ferroelectric memristor. This device functions as a voltage controllable variable resistor, which can be applied to a synapse weight. However, its conductance shows hysteresis characteristics and dispersion to the input voltage. Therefore, the conductance values vary according to the history of the height and the width of the applied pulse voltage. Due to the difficulty of controlling the accurate conductance, it is not easy to apply the back-propagation learning algorithm to the neural network hardware having memristor synapses. To solve this problem, we proposed and simulated a learning operation procedure as follows. Employing a weight perturbation technique, we derived the error change. When the error reduced, the next pulse voltage was updated according to the back-propagation learning algorithm. If the error increased the amplitude of the next voltage pulse was set in such way as to cause similar memristor conductance but in the opposite voltage scanning direction. By this operation, we could eliminate the hysteresis and confirmed that the simulation of the learning operation converged. We also adopted conductance dispersion numerically in the simulation. We examined the probability that the error decreased to a designated value within a predetermined loop number. The ferroelectric has the characteristics that the magnitude of polarization does not become smaller when voltages having the same polarity are applied. These characteristics greatly improved the probability even if the learning rate was small, if the magnitude of the dispersion is adequate. Because the dispersion of analog circuit elements is inevitable, this learning operation procedure is useful for analog neural network hardware.

Citation: Ueda M, Nishitani Y, Kaneko Y, Omote A (2014) Back-Propagation Operation for Analog Neural Network Hardware with Synapse Components Having Hysteresis Characteristics. PLoS ONE 9(11): e112659. doi:10.1371/journal.pone.0112659

Editor: Mark D. McDonnell, University of South Australia, Australia

Received: June 5, 2014; **Accepted:** October 20, 2014; **Published:** November 13, 2014

Copyright: © 2014 Ueda et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability: The authors confirm that all data underlying the findings are fully available without restriction. All relevant data are within the paper.

Funding: The authors have no support or funding to report.

Competing Interests: The authors are employees of Panasonic Corporation. This does not alter the authors' adherence to all the PLOS ONE policies on sharing data and materials.

* Email: ueda.michihito@jp.panasonic.com

Introduction

The artificial neural network (ANN) is receiving research interest, for example, due to deep learning approaches that are improving recognition rates in benchmark classification problems [1,2]. There have been studies on large-scale digital processing built upon the conventional CPUs and GPUs [3]. However, built only with digital circuits [4], the ANN hardware requires a large volume of memory. It is true that the algorithm improvement can reduce the memory size, but cannot solve the fundamental problem. A hardware-level solution must be proposed.

One of the solutions is introducing a neuromorphic device. To realize ANN hardware, the circuit element for synapse function is important because the number of synapse elements is much larger than that of neuron elements. One of the candidates for this synapse element is a memristor [5,6]. Because the conductance of the memristor depends on the history of the applied voltage, it can realize the synapse function [7,8]. The memristor-based memories can achieve a very high integration density of 100 Gbit/cm², a few times higher than flash memory technologies [9]. These unique properties make it a promising device for massively parallel, large-

scale neuromorphic systems [7,10]. Hu *et al.* have also reported the potential of a memristor crossbar array that functions as an associative memory [11].

We have also examined the synapse function using a ferroelectric memristor (FeMEM) [12,13]. Because FeMEM could be operated at a 60 nm channel length [14], high density integration of FeMEM synapse device can be expected. We demonstrated the conductance change according to the biologically inspired learning method of spike-timing-dependent synaptic plasticity (STDP) [15,16]. As the FeMEM has three terminals, concurrent learning can be realized. We constructed an analog circuit with FeMEM synapses for a Hopfield neural network, and by using this STDP learning method, we demonstrated the learning and recalling of patterns [17,18].

To realize generic ANN hardware, we should adapt the learning method to a back-propagation (BP) algorithm [19]. Ishii *et al.* reported hardware BP learning for neuron metal-oxide-semiconductor (MOS) neural networks [20]. However, the neuron MOS did not have non-volatile memory to store the learned synapse weight.

By applying a memristor as a multivalued memory, many researchers have reported ANN hardware having memristor synapses [6–8,11,17,18,21]. However, a memristor has hysteresis characteristics of input voltage and conductance. The conductance values vary according to the history of the applied voltage and its width. These characteristics make it difficult to control its conductance to the desired value. Therefore, it is not easy to apply the BP learning algorithm to ANN hardware having memristor synapses. The purpose of this paper is to develop a simple procedure for the BP learning operation, which can be applied to analog ANN hardware with synapse devices having hysteresis and variability.

Analog Neural Network Hardware with FeMEMs

1. Feed-forward neural network

Figure 1 shows the analyzed feed-forward neural network structure. This structure has two inputs in the input layer, three neurons in the hidden layer, and one neuron in the output-layer. A neuron has multiple synapses. A fundamental calculation for neural networks is the product sum operation described by

$$M(i) = f\left(\sum_{j=1}^3 w_m(i,j) S_{in}(j)\right), \quad (1)$$

$$S_{out} = f\left(\sum_{i=1}^4 w_o(i) M(i)\right), \quad (2)$$

where $M(i)$ is the output from hidden-layer neurons, S_{out} is the output from the output-layer neuron, $S_{in}(1)$ and $S_{in}(2)$ are two inputs, $w_m(i, j)$ and $w_o(i)$ are the synapse weights of the hidden-layer neurons and the output-layer neurons, respectively. The function $f(-)$ is a threshold function; a sigmoidal function is

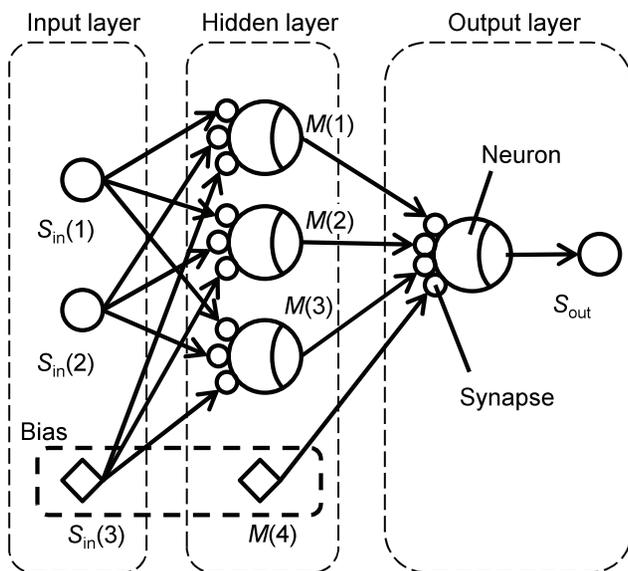


Figure 1. Analyzed feed forward neural network. The structure has two inputs in the input layer, three neurons in the hidden layer, and one neuron in the output layer. $S_{in}(1)$ and $S_{in}(2)$ are the inputs. $M(1)$ – $M(3)$ and S_{out} are the output from hidden and output layers, respectively. $S_{in}(3)$ and $M(4)$ are the bias inputs. doi:10.1371/journal.pone.0112659.g001

frequently used. In Figure 1, both $S_{in}(3)$ and $M(4)$ are bias inputs, and their values are unity.

2. ANN hardware

To realize an analog neuron device, we examined a circuit based on an operational amplifier (op-amp) adder circuit. Using FeMEMs and an op-amp, a neuron circuit was constructed as shown in Figure 2(a). R_F is a fixed resistance, whose conductance is G_R . To achieve a synapse function using an FeMEM, the synaptic circuit modules were devised that consist of inhibitory/excitatory synapse pairs. As the op-amp adder circuit is an inverting amplifier circuit, the inhibitory pairs receive raw input directly and the excitatory ones receive inverted copies of the raw input voltage via a unity gain inverting amplifier. Although this synapse circuit construction needs two FeMEMs, a highly functional neuron circuit can be realized, because the modulation of synapse weight is easier to control individually with two FeMEMs. Here, we denote the channel conductance of FeMEM as G_F . Also, we denote G_F for the excitatory synapse as $G_E(i)$ and G_F for the inhibitory synapse as $G_I(i)$. The sum of amplified voltages, or the inner potential (u), is calculated as

$$u = \frac{1}{G_R} \sum_{i=1}^{N_{in}} V_{in}(i) \{G_E(i) - G_I(i)\}, \quad (3)$$

where N_{in} is the total number of inputs, $V_{in}(i)$ is the input voltage. The non-linear output voltage (V_{out}) of the op-amp is

$$V_{out} = \begin{cases} 1 \text{ V} & (u < -1 \text{ V}) \\ -u & (-1 \text{ V} \leq u \leq 1 \text{ V}) \\ -1 \text{ V} & (u > 1 \text{ V}) \end{cases}. \quad (4)$$

As this circuit is an inverting amplifier circuit, the plus and minus signs of u are reversed for the output voltage. Thus, the input voltage for excitatory FeMEM is reversed by using an inversion circuit. Using G_F , the synapse weight (w) can be calculated as G_F/G_R . We use the output of the op-amp as that of a neuron circuit for convenience in constructing the circuit, although, in general neural networks, the output of a neuron is calculated using a threshold function such as a sigmoidal function. Figure 2(b) compares a sigmoidal function and the op-amp output. The change in the op-amp output is linear in the voltage range of the power supply and is constant out of the range. Though the linear property of the op-amp may degrade the learning ability, the sigmoidal function and op-amp output have similar trends.

Preparation of ANN Hardware and Proposal of Learning Operation

1. Structure and procedure for preparation of the FeMEM

We fabricated a FeMEM structure based on insights gained in previous studies [12–14]. As shown in Figure 3(a) (b), the FeMEM consists of a semiconductor film of ZnO, a ferroelectric film of Pb(Zr,Ti)O₃ (PZT), and a bottom gate electrode of SrRuO₃ (SRO). All the layers of ZnO/PZT/SRO were epitaxially grown over a SrTiO₃ (STO) substrate by pulsed laser deposition. Pt/Ti electrodes were used for the source and drain contacts to the ZnO film. The fabricated FeMEM showed electron gas accumulation and complete depletion switching operation due to reversal of the ferroelectric polarization. The channel conductance (G_F)–gate

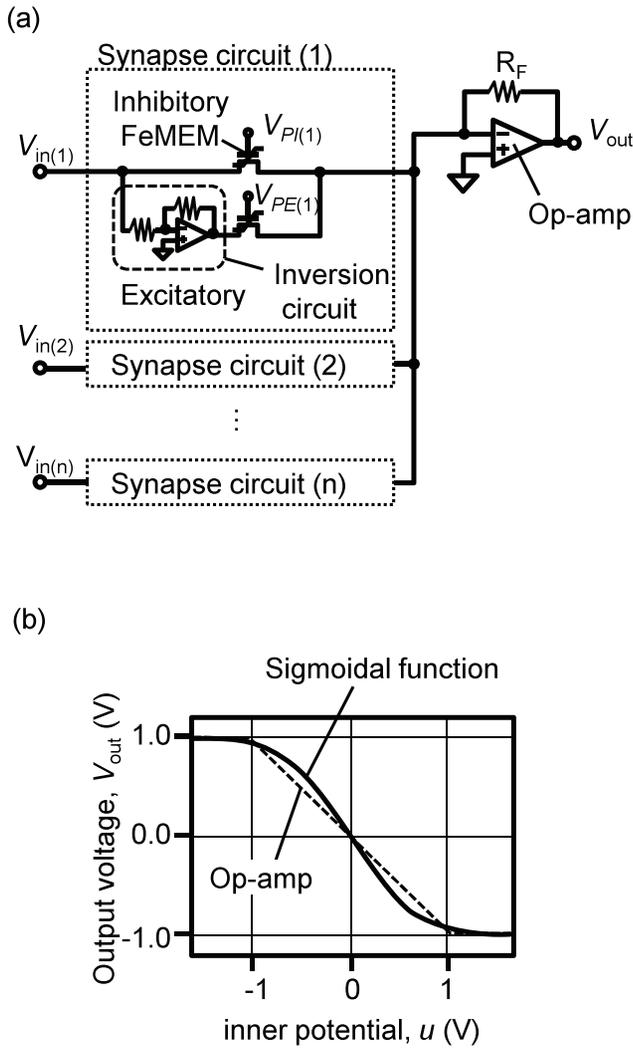


Figure 2. Schematic of a neuron circuit and its input-output characteristics. The neuron circuit is based on an op-amp adder circuit as shown in (a). Synapse circuits are constructed with a FeMEM. To realize positive and negative synapse weights, we adopted excitatory and inhibitory synapses. The inner potential (u) is calculated according to (3). The relation between u and output voltage (V_{out}) of the op-amp resembles the input-output characteristics of a sigmoidal function as shown in (b).

doi:10.1371/journal.pone.0112659.g002

voltage (V_G) characteristics of the FeMEM are shown in Figure 3(c). The G_F - V_G characteristics were measured using a semiconductor parametric analyzer (Agilent 4155C) under the condition of long integration time. By measuring the drain current under the condition of drain voltage = 0.1 V, G_F was calculated. The drain voltage was set to be low so as not to change the polarization of the ferroelectric. The figure shows counterclockwise hysteresis loops corresponding to the switching of ferroelectric polarization. The conductance at $V_G = 0$ V changed according to the history of applied V_G and could thus take multiple values. It was confirmed that there was no notable degradation of conductance over 10^5 s [12]. These characteristics allowed the construction of an analog ANN circuit with synapse elements using the FeMEM [15–18,21].

2. Electrical characteristics of the synapse circuit

We examined the performance of the basic neuron circuit. The experimental setup used to evaluate the relation of the pulse voltage (V_P) and the conductance of the FeMEM is shown in Figure 4(a). The devices we use in this experiment have been tested before and were found to exhibit good non-volatility characteristics [12]. The pulse width of V_P was set to 1 ms. To enhance the conductance repeatability, the conductance was measured after applying a reset pulse (V_R). $V_R = -2$ V when $V_P > 0$ and $V_R = 3$ V when $V_P < 0$. V_P was first increased from 0 to 3 V in 0.2 V steps and then reduced from 0 to -2 V in -0.2 V steps. In the same manner as G_F - V_G measurement, the drain current was measured under the condition of drain voltage = 0.1 so as not to change the polarization of the ferroelectric.

This scanning operation was performed 300 times. From the measured V_{out} , G_F was calculated according to

$$G_F = -G_R V_{out} / V_{in} \quad (5)$$

Because V_R and V_P are pulse voltages, the voltages are applied only during the weight update. This enables the reduction of the number of voltage sources as the pulse voltages can be applied by switching a voltage source. Moreover, the power consumption for maintaining the synapse weight is zero.

The average and standard deviation of calculated conductance are shown in Figure 4(b). Smooth counterclockwise characteristics were observed. The conductance change was in the range of 0.5×10^{-6} – 40×10^{-6} S for the investigated V_P range.

To analyze a learning operation, as an alternative approach, we prepared a numerical model of the FeMEM conductance by fitting experimental data. We fitted the average conductance with sigmoidal functions that are commonly used in modeling ferroelectric functions [22]. We manually fitted the two curves of increasing voltage and decreasing voltage and derived the equation.

$$G_F = G_{min} + \frac{G_{max} - G_{min}}{1 + \exp[-\alpha(V_P - \theta)]}, \quad (6)$$

where $\alpha = 3 \text{ V}^{-1}$, $G_{max} = 45 \times 10^{-6}$ S, $G_{min} = 0.5 \times 10^{-6}$ S, $\theta (= \theta_1) = 2.3$ V (for increasing V_P) and $\theta (= \theta_2) = -0.6$ V (for decreasing V_P). The fitting curves are shown in Figure 4(b) as broken lines. Approximate characteristics were expressed, regardless of there being few parameters.

3. Proposal of BP operation of hysteresis synapse devices

BP is the most widely applied learning methods for training an ANN. When a synapse weight (w) is changed, the outputs of the ANN also change, which is known as weight perturbation [23]. The synapse weight w is updated according to

$$w_{new} = w_{old} - \eta \Delta E / \Delta w, \quad (7)$$

where η is the learning rate, w_{new} and w_{old} are the synapse weights after and before the update, respectively. The square error (E) is calculated according to

$$E = (T_{out} - S_{out})^2 / 2, \quad (8)$$

where T_{out} is a target output. Because the ANN in this study has

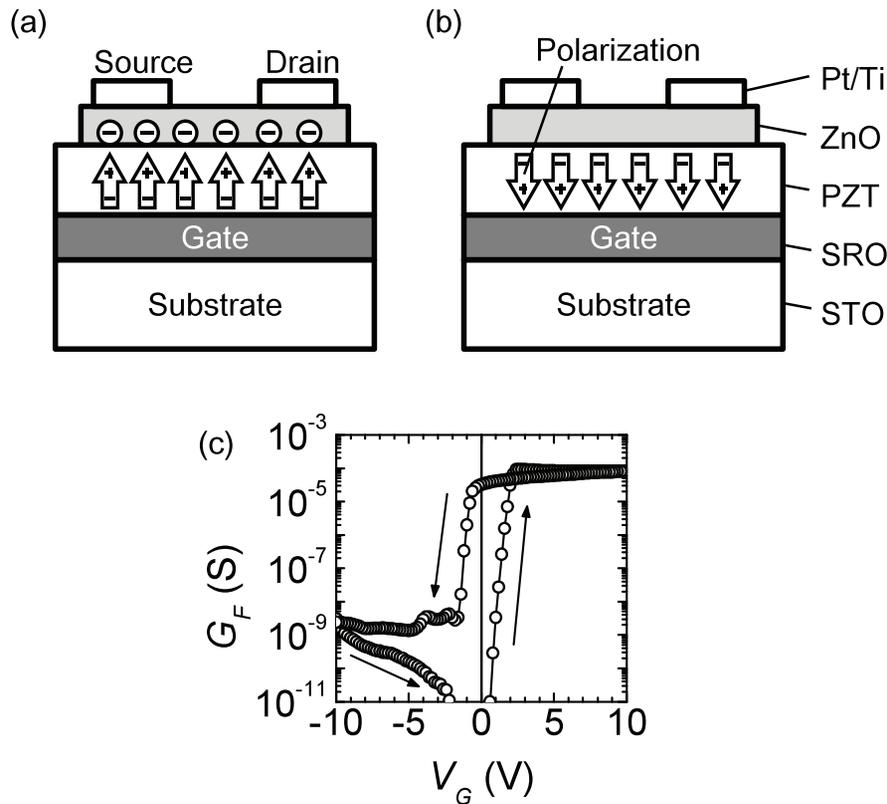


Figure 3. Schematics of FeMEM and its electrical properties. The fabricated FeMEM showed (a) electron gas accumulation and (b) complete depletion switching operation due to the reversal of ferroelectric polarization. (c) G_F - V_G characteristics showed the counterclockwise hysteresis loop corresponding to the switching of ferroelectric polarization. doi:10.1371/journal.pone.0112659.g003

only one output, E can be calculated simply according to (8). ΔE is the difference between the square errors after and before the update and is calculated according to

$$\Delta E^n = E_2^n - E_1^n = \left(S_{\text{out}2}^n - S_{\text{out}1}^n \right) \left(S_{\text{out}2}^n + S_{\text{out}1}^n - 2T_{\text{out}}^n \right) / 2 \quad (9)$$

where the subscript 1 and 2 indicates before and after the update, and the superscript n indicates the input pattern number defined in table 1.

As the synapse weight (w) of the analog ANN hardware in this paper is calculated as $w = G_F / G_R$, w is proportional to G_F . Because G_F is a function of V_P as shown in (6), we simply updated V_P according to

$$V_{P(\text{new})} = V_{P(\text{old})} - \eta \Delta E / \Delta V \quad (10)$$

where $V_{P(\text{new})}$ and $V_{P(\text{old})}$ are V_P s after and before the update, respectively, and ΔV is the minute V_P change to obtain the error difference.

Moreover, to eliminate the effect of the hysteresis of the conductance, we applied the following procedures. The detailed flowchart is shown in Figure 5. In this flowchart, V_{PE} and V_{PI} are V_P s for the excitatory and inhibitory synapses, respectively, whose values are stored in an external memory. θ is the designated threshold value to exit from the learning procedure. E_{sum} is the sum of the square errors for all target output values and is calculated according to:

$$E_{\text{sum}} = \sum_{n=1}^4 \Delta E^n \quad (11)$$

The learning procedures are as follows:

- Select a synapse to update. In this paper, we started from output-layer.
- First, the FeMEMs for excitatory synapses are updated; $V_{P(\text{old})} = V_{PE}$.
- Error E_1 at a point is calculated according to the outputs from ANN hardware and target output values.
- Slightly larger V_P in amplitude than the stored previous $V_{P(\text{old})}$ is applied to the FeMEM constructing the target synapse. That is, when $\Delta V > 0$, $V_{P(\text{old})} + \Delta V$ is applied in the case of increasing V_P , and $V_{P(\text{old})} - \Delta V$ is applied in the case of decreasing V_P .
- Error E_2 at a point is calculated in the same manner as step C.
- If $E_2 \leq E_1$, according to (10), $V_{P(\text{new})}$ is updated and stored in an external memory. In this case, V_R is not applied. We term this operation “non-inverting operation”.
- If $E_2 > E_1$, V_R is applied according to V_P polarity as explained in Figure 4(b). Subsequently, V_P of the same conductance on another curve in Figure 4(b) is updated and stored; *i.e.*, $V_{P(\text{new})} = V_P - (\theta_1 - \theta_2)$ if V_P is increasing and

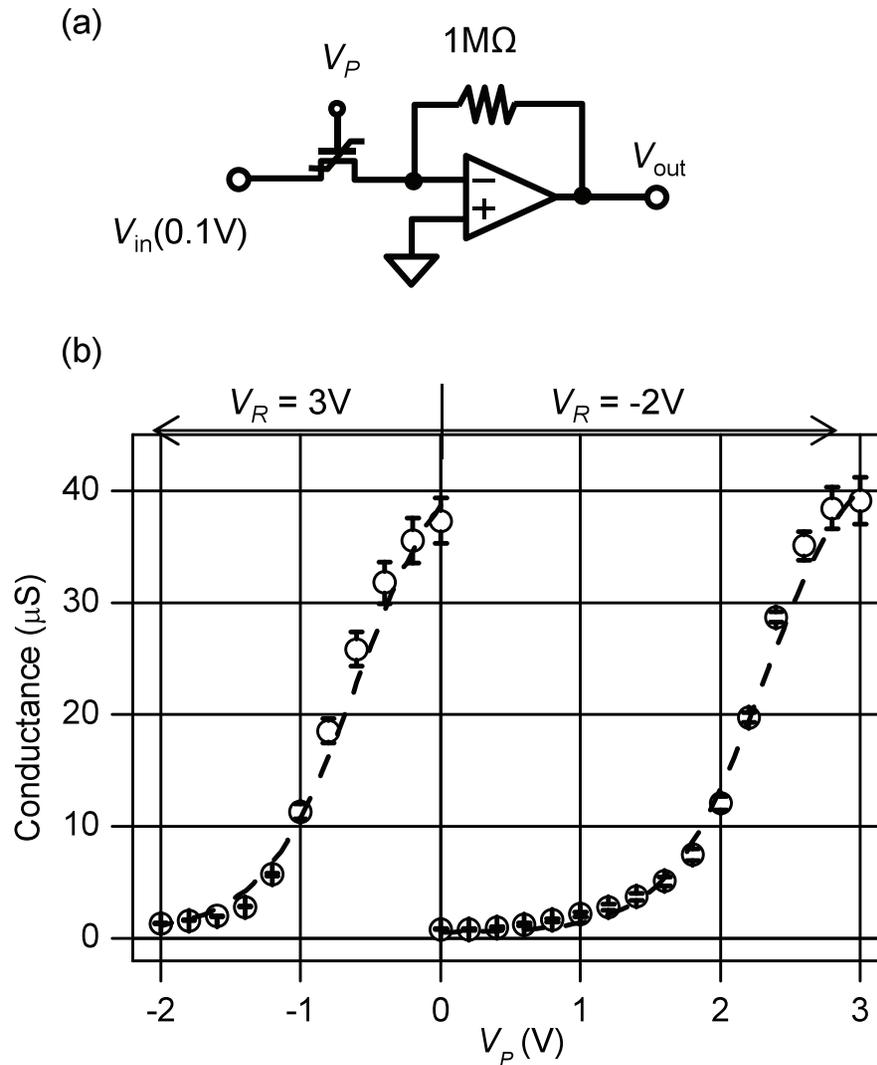


Figure 4. Schematics of measurement setup and calculated conductance. With the measurement setup shown in (a), the conductance of the FeMEM can be calculated from the measured output voltage (V_{out}) from the op-amp when input voltage $V_{in} = 0.1$ V. After applying a reset pulse (V_R) and a write pulse (V_P) to the gate electrode of the FeMEM, V_{out} is measured. The calculated conductance is shown in (b). The open circles indicate the average values and the error bars indicate the standard deviation over 300 scans.
doi:10.1371/journal.pone.0112659.g004

$V_{P(new)} = V_P + (\theta_1 - \theta_2)$ if V_P is decreasing. We term this operation “inverting operation”.

- H. For the FeMEMs for inhibitory synapses, by substituting $V_{P(old)}$ to V_{PI} , V_{Ps} are updated in the same manner as steps C to G.
- I. Steps A to H are repeated for all synapses in the network.

- J. If E_{sum} is larger than Θ , then the process returns to step A, else the learning procedure is finished.

At the step G, V_P is switched and jumps from one curve to another in Figure 4(b). The reset pulse V_R and this V_P jump eliminate the hysteresis of $\theta_1 - \theta_2$.

Table 1. Definition of the input pattern number n .

n	$S_{in(1)}$	$S_{in(2)}$
1	1 V	1 V
2	1 V	-1 V
3	-1 V	1 V
4	-1 V	-1 V

doi:10.1371/journal.pone.0112659.t001

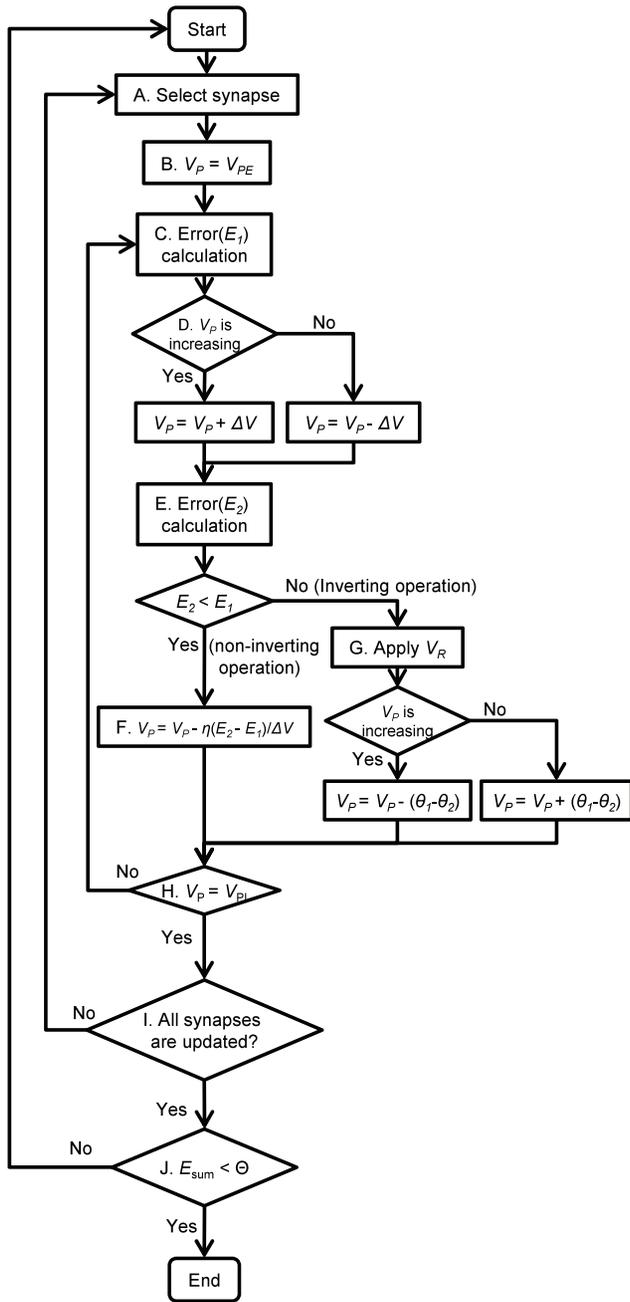


Figure 5. Flowchart of back-propagation learning operation. The write pulses (V_p) for the excitatory and inhibitory synapses are defined as V_{PE} and V_{PI} , respectively. V_p is slightly changed (ΔV) according to the V_p direction and the error change is calculated. When the error increases ($E_2 > E_1$), after applying V_R at step G, V_p is switched and jumps from one curve to another in Figure 4(b). This V_p jump eliminates the hysteresis of $\theta_1 - \theta_2$. doi:10.1371/journal.pone.0112659.g005

Results and Discussions

1. Learning of the boolean logic of exclusive OR

To evaluate the proposed learning operation, we numerically analyzed the learning process of the Boolean logic of exclusive OR using (10). The high and low signals were set to 1 and -1 V, respectively. In this analysis, $\Delta V = 10$ mV, $\eta = 0.01$ V^{-2} and $G_R = 10^{-5}$ S. Initial values of G_F were

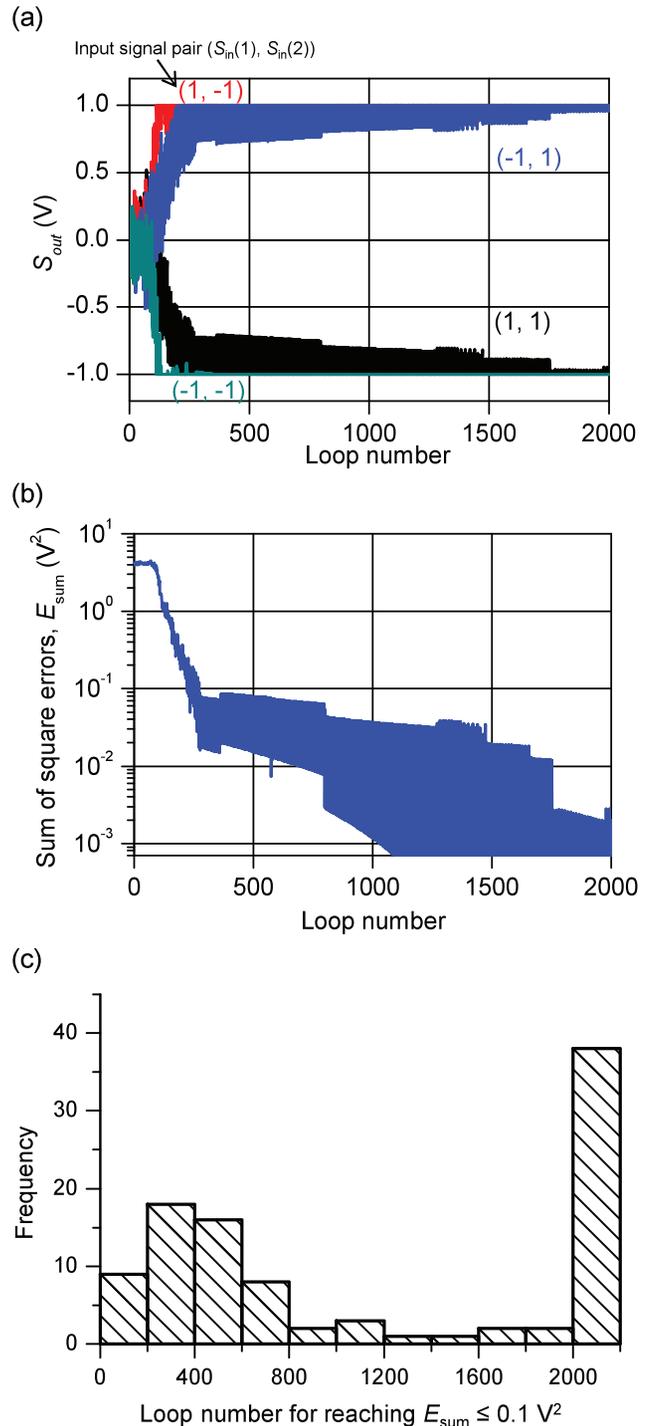


Figure 6. Typical example of the learning operation under the condition of $\eta = 0.01$ V^{-2} . Typical example of the learning operation under the condition of the learning rate (η) = 0.01 V^{-2} . (a) output from the output-layer neuron (S_{out}) evolution for four input signal pairs. (b) the sum of the square errors (E_{sum}) evolution. (c) Histogram of loop number for reaching $E_{sum} \leq 0.1$ V^2 . We denote the probability for reaching $E_{sum} \leq 0.1$ V^2 within 2000 loops as P_C . In this case, P_C was about 60%. doi:10.1371/journal.pone.0112659.g006

randomly set within $[0.9 \times 10^{-6}, 1.1 \times 10^{-6}]$ S. The learning operation for the output layer was first executed and then the learning operation for the hidden layer. The results are shown

in Figure 6. One loop involves the update of all synapses in output and hidden layers.

S_{out} values fluctuated until about 200 loops; however, the values gradually became correct afterward (Figure 6(a)). E_{sum} gradually decreased from 200 loops, and the learning operation successfully converged (Figure 6(b)). By changing the initial values of G_F randomly, we simulated 100 learning processes and examined the loop number required for reaching $E_{sum} \leq 0.1 \text{ V}^2$. As shown in Figure 6(c), the loop number for reaching $E_{sum} \leq 0.1 \text{ V}^2$ with the highest frequency was 200–400 bin. However, there were cases in which E_{sum} was larger than 0.1 V^2 after more than 2000 loops. Here we denote the probability for reaching $E_{sum} \leq 0.1 \text{ V}^2$ within 2000 loops as P_C . In this case, P_C was about 60%.

2. Adoption of conductance dispersion

In the Section 4.1, a simulation was carried out for the conductance characteristics of the FeMEM using (6). However, as seen from Figure 4(b), the conductance showed dispersion. In this Section, we adopt this dispersion numerically to simulate under a more realistic condition. From the results in Figure 4(b), the coefficient of variation (CV), which is calculated by dividing the standard deviation by the mean, was plotted as shown in Figure 7.

The results show that the CV is less than 0.1 for all values of conductance. Therefore, when we calculate G_F for applied V_P , we introduced random dispersion so that the $CV = 0.1$. The conductance G_F' , which involves the dispersion, is calculated according to

$$G_F' = G_F + \xi, \quad (12)$$

where ξ indicates the Gaussian dispersion, which is calculated by Box–Muller method.

Here, we introduce the properties of ferroelectric material to this ξ . Under the condition that voltage pulses have the same polarity and sufficient pulse widths, the polarization of the ferroelectric changes only when the maximum amplitude voltage

in its history is applied. In this experiment, the pulse width of V_P was set to 1 ms, which is sufficiently wide because the switching time of this device is less than $1 \mu\text{s}$ [13].

In the inverting operation, because the polarity changes, G_F' changes according to (12). However, in the non-inverting operation, the G_F' changes only when the maximum amplitude voltage is applied according to the properties of ferroelectric. As a result, G_F' never decreases in cases of $V_P > 0$ and never increases in cases of $V_P < 0$. In the following, we term this “restriction effect”. In the simulation, this effect was realized by setting $\xi = 0$ in the case that ($\xi < 0$ and $V_P > 0$) or ($\xi > 0$ and $V_P < 0$).

When the error is decreasing ($E_2 < E_1$), because the non-inverting operation is chosen, if ξ is not too large, the error hardly increases by G_F' . Needless to say, if ξ is too large, because G_F' jumped over the adequate conductance value, the error increases. It should be noted that the restriction effect enhanced the correct G_F' change.

Taking this restriction effect into consideration, we simulated the learning process under the condition of $CV > 0$. The simulation results are shown in Figure 8. Though, both S_{out} and E_{sum} showed large fluctuation, E_{sum} rapidly decreased from 400 loops and fallen below the designated value at about 450 loops. In Figure 8(b), this point is indicated as “Reaching point”. However, after that, E_{sum} rapidly increased again. These results are very different from those of $CV = 0$ in Figure 6. When $CV = 0$, E_{sum} decreased gradually from 200 loops and, after that, continued decreasing. When $CV = 0.1$, E_{sum} fluctuated in large scale, however, E_{sum} was not always large but fall down again below 10^{-3} V^2 . The results seemed not to diverge.

Finally, to clarify the effect of the conductance dispersion, we analyzed the relation between P_C and η changing CV values. The results are shown in Figure 9.

When $\eta \gg 0.01 \text{ V}^{-2}$, regardless of CV value, because V_P (and G_F) changed too large according to (12), G_F jumped over the adequate value. As a result, P_C became small.

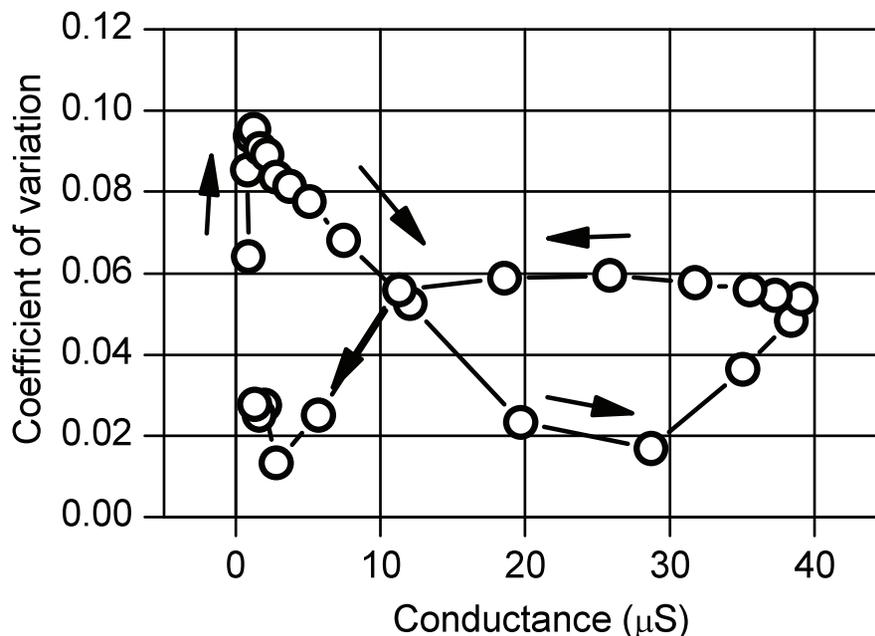


Figure 7. Coefficient of variation of conductance. Coefficient of variation (CV) of conductance is calculated by dividing the standard deviation by the mean. The results show that the CV is less than 0.1 for all values of conductance. doi:10.1371/journal.pone.0112659.g007

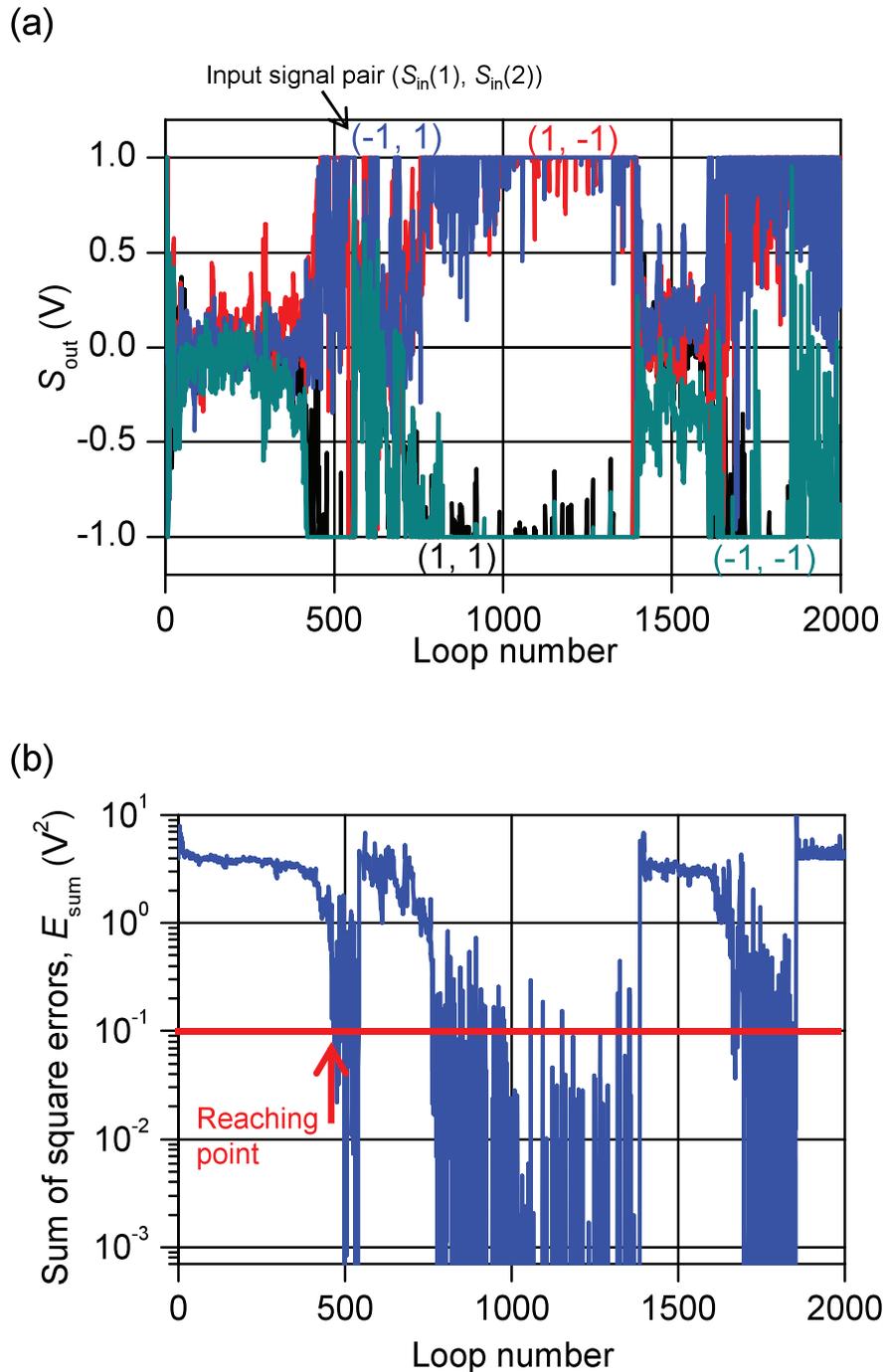


Figure 8. Typical example of the learning operation under the condition of $CV=0.1$ and $\eta=0.01 \text{ V}^{-2}$. Typical example of the learning operation under the condition of the coefficient of variation (CV) = 0.1 and the learning rate (η) = 0.01 V^{-2} . (a) output from the output-layer neuron (S_{out}) evolution for the four input signal pairs. (b) the sum of the square errors (E_{sum}) evolution. “Reaching point” indicates the loop number at which the E_{sum} is smaller than 0.1 V^2 for the first time in the operation.
doi:10.1371/journal.pone.0112659.g008

When $\eta \ll 0.01 \text{ V}^{-2}$, under the condition that $CV=0$, V_P (and G_F) changed so small that E_{sum} hardly changed. Consequently, P_C became small because P_C is defined as the probability of $E_{sum} \leq 0.1 \text{ V}^2$ within 2000 loops. In this case, the maximum P_C of about 60% was obtained around $\eta = 0.01 \text{ V}^{-2}$. On the other hand, under the condition of $CV > 0$, though V_P hardly changed, the learning operation progressed because the G_F' changed. Moreover, in almost cases, E_{sum} was expected to decrease by the restriction

effect in case of non-inverting operation. Although it was true that there was a possibility that E_{sum} increased in case of inverting operation, it was shown that high P_C was realized in a large region of $\eta \leq 0.01 \text{ V}^{-2}$, especially $CV \sim 0.1$. When CV is too large ($CV \geq 0.3$), as explained above, because G_F' jumped over the adequate value, the error increased and P_C became small regardless of η value.

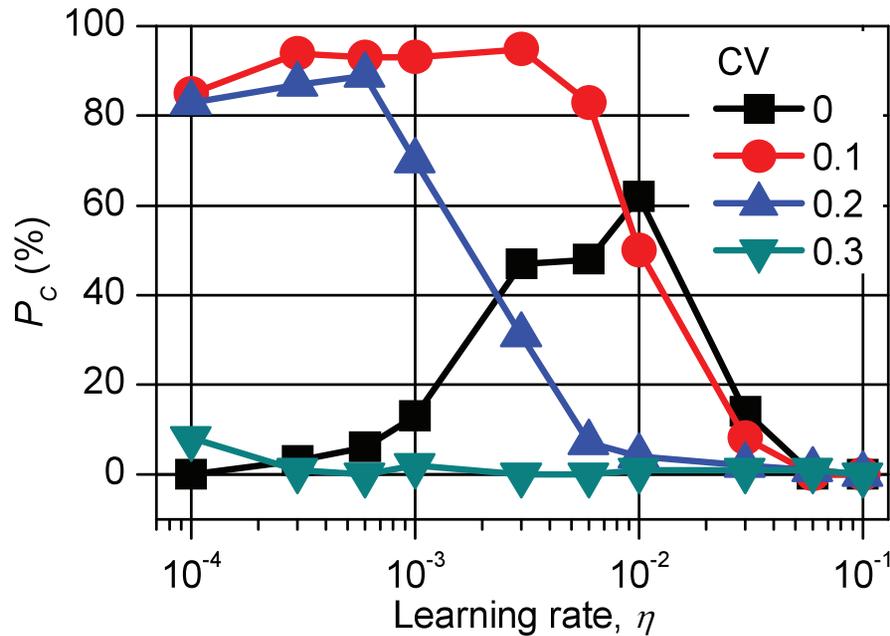


Figure 9. Relation between P_C and η under the condition of coefficient of variation = 0–0.3. P_C is the probability for reaching sum of the square errors (E_{sum}) $\leq 0.1 \text{ V}^2$ within 2000 loops, and η is the learning rate. When the coefficient of variation (CV) is appropriate, the pulse voltage changes adequately even though $\eta = 0 \text{ V}^{-2}$. These results show that, in wide regions of the CV, high P_C is realized. When $\eta \gg 0.01 \text{ V}^{-2}$, the pulse voltage (V_P) changed so large according to (10) that the conductance of FeMEM (G_F) also changed large and that P_C became small regardless of CV. When $\eta \ll 0.01 \text{ V}^{-2}$, V_P hardly changed in case of CV=0. However, in case of CV=0.1, the learning operation progressed because the G_F is not so small according to (12). Moreover, in almost cases, E_{sum} was expected to decrease by the restriction effect in non-inverting operation. doi:10.1371/journal.pone.0112659.g009

The CV is a difficult parameter to control. When the absolute value of reset voltage ($|V_R|$) is higher, the repeatability of conductance improves because the ferroelectric polarization is along the major loop, whereas when the $|V_R|$ is lower, because the ferroelectric polarization is along the minor loop, the dispersion of conductance increases. Thus, rough control of the CV is possible by changing the reset voltage value. Because P_C is high in a comparatively wide region of the CV, high P_C can be achieved by controlling the reset voltage.

As for neural networks, it is commonly known that noise assists in the escape from local minima [24]. Conversely, for analog ANN hardware, noise is harmful to learning because voltages cannot be controlled strictly. As for FeMEM, however, we found that appropriate dispersion realized large P_C value.

The proposed operation procedure is simple and easy to implement in hardware yet is capable of eliminating the effect of hysteresis and is robust against the dispersion of conductance. Another type of memristor also displays the hysteresis [7,8]. By expressing the relation between the conductance and the applied voltage as equations and analyzing the CV, our approach can be also applicable to such memristors if they also exhibit restriction effect.

Conclusions

A BP learning operation was studied for analog artificial neural network (ANN) hardware having a ferroelectric memristor (FeMEM) synapse. The synapse weight was expressed by the channel conductance (G_F) of the FeMEM. After applying a reset pulse, by changing the height of the pulse voltage (V_P), smooth counterclockwise characteristics of G_F - V_P were observed.

To eliminate the effect of hysteresis of the conductance, we proposed a learning operation, by which G_F always traveled on either curve of two G_F - V_P relations. By this operation, because G_F traveled practically along a continuous function, we confirmed that the simulation of the learning operation converged.

The measured G_F had a coefficient of variation up to 0.1. Therefore, we adopted conductance dispersion numerically in the simulation. As a result, the dispersion introduced large fluctuation in the converging process; however, the probability (P_C) for reaching $E_{sum} \leq 0.1 \text{ V}^2$ within 2000 loops was not so poor. Moreover, when the learning rate was smaller than 0.01, P_C greatly improved to 85%. These results were obtained by the properties of ferroelectric. When V_P was not inverted, the dispersion affects only in the direction of decreasing error. Dispersion is not a controllable parameter but a characteristic of the FeMEM; however, it can be roughly changed by the reset voltage.

The proposed operation procedure is simple and easy to implement in hardware. Considering the inevitability of the dispersion of analog circuit elements, this operating procedure is useful for analog ANN hardware. As the scale of ANN processing is increasingly growing, analog ANN hardware is a promising candidate for effective calculation and will play an important role in energy saving for large-scale ANNs.

Author Contributions

Conceived and designed the experiments: MU AO. Performed the experiments: MU YK. Analyzed the data: MU YN. Contributed reagents/materials/analysis tools: YK. Contributed to the writing of the manuscript: MU.

References

- Hinton GE, Osindero S, Teh Y-W (2006) A fast learning algorithm for deep belief nets. *Neural Comput* 18: 1527–1554.
- Le Q V (2013) Building high-level features using large scale unsupervised learning. *IEEE Int Conf on Acoustics, Speech, and Sig Proc (ICASSP)*: 8595–8598.
- Coates A, Huval B, Wang T, Wu DJ, Ng AY, et al. (2013) Deep learning with COTS HPC systems. *Proc 30th Int Conf Mach Learn*: 1337–1345.
- Partzsch J, Schüffny R (2011) Analyzing the scaling of connectivity in neuromorphic hardware and in models of neural networks. *IEEE Trans Neural Netw* 22: 919–935.
- Strukov DB, Snider GS, Stewart DR, Williams RS (2008) The missing memristor found. *Nature* 453: 80–83.
- Alibart F, Zamanidoost E, Strukov DB (2013) Pattern classification by memristive crossbar circuits using ex situ and in situ training. *Nat Commun* 4: 2072.
- Jo SH, Chang T, Ebong I, Bhadviya BB, Mazumder P, et al. (2010) Nanoscale memristor device as synapse in neuromorphic systems. *Nano Lett* 10: 1297–1301.
- Kuzum D, Jeyasingh RGD, Lee B, Wong HSP (2012) Nanoelectronic programmable synapses based on phase change materials for brain-inspired computing. *Nano Lett* 12: 2179–2186.
- Ho Y, Huang G, Li P (2009) Nonvolatile memristor memory: device characteristics and design implications. *IEEE/ACM International Conference on Computer-Aided Design (ICCAD)*: 485–490.
- Xia Q, Robinett W, Cumbie MW, Banerjee N, Cardinali TJ, et al. (2009) Memristor-CMOS hybrid integrated circuits for reconfigurable logic. *Nano Lett* 9: 3640–3645.
- Hu M, Li H, Chen Y, Wu Q, Rose GS, et al. (2014) Memristor Crossbar-Based Neuromorphic Computing System: A Case Study. *IEEE Trans Neural Networks Learn Syst*. Available: <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6709674>.
- Kato Y, Kaneko Y, Tanaka H, Shimada Y (2008) Nonvolatile Memory Using Epitaxially Grown Composite-Oxide-Film Technology. *Jpn J Appl Phys* 47: 2719–2724.
- Kaneko Y, Nishitani Y, Tanaka H, Ueda M, Kato Y, et al. (2011) Correlated motion dynamics of electron channels and domain walls in a ferroelectric-gate thin-film transistor consisting of a ZnO/Pb(Zr,Ti)O₃ stacked structure. *J Appl Phys* 110: 084106.
- Kaneko Y, Nishitani Y, Ueda M, Tokumitsu E, Fujii E (2011) A 60 nm channel length ferroelectric-gate field-effect transistor capable of fast switching and multilevel programming. *Appl Phys Lett* 99.
- Nishitani Y, Kaneko Y, Ueda M, Morie T, Fujii E (2012) Three-terminal ferroelectric synapse device with concurrent learning function for artificial neural networks. *J Appl Phys* 111.
- Nishitani Y, Kaneko Y, Ueda M, Fujii E, Tsujimura A (2013) Dynamic Observation of Brain-Like Learning in a Ferroelectric Synapse Device. *Jpn J Appl Phys* 52: 04CE06.
- Kaneko Y, Nishitani Y, Ueda M, Tsujimura A (2013) Neural network based on a three-terminal ferroelectric memristor to enable on-chip pattern recognition. *Symposia on VLSI Technology and Circuits*: T238–T239.
- Kaneko Y, Nishitani Y, Ueda M (2014) Ferroelectric Artificial Synapses for Recognition of a Multishaded Image. *IEEE Trans Electron Devices* 61: 2827–2833. Available: <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6848780>.
- Bryson AE, Ho Y-C (1969) *Applied Optimal Control: Optimization, Estimation and Control* Blaisdell Publishing Company. p. 481.
- Ishii H, Shibata T, Kosaka H, Ohmi T (1992) Hardware-backpropagation learning of neuron MOS neural networks. *IEEE International Electron Devices Meeting*: 435–438.
- Ueda M, Kaneko Y, Nishitani Y, Fujii E (2011) A neural network circuit using persistent interfacial conducting heterostructures. *J Appl Phys* 110: 086104.
- Miller SL, Nasby RD, Schwank JR, Rodgers MS, Dressendorfer P V (1990) Device modeling of ferroelectric capacitors. *J Appl Phys* 68: 6463. doi:10.1063/1.346845.
- Jabri M, Flower B (1992) Weight perturbation: An optimal architecture and learning technique for analog VLSI feedforward and recurrent multilayer networks. *IEEE Trans Neural Networks* 3: 154–157.
- Wang C, Principe JC (1999) Training neural networks with additive noise in the desired signal. *IEEE Trans Neural Networks* 10: 1511–1517.