A PWM Analog Memory Programming Circuit for Floating-Gate MOSFETs with 75-μs Programming Time and 11-bit Updating Resolution

Shigeo Kinoshita, Takashi Morie, Makoto Nagata, and Atsushi Iwata

Abstract—**This paper describes a programming circuit for analog memory using pulsewidth modulation (PWM) signals and the circuit performance obtained from measurements using a floating-gate EEPROM device. This programming circuit attains both high programming speed and high precision. We fabricated the programming circuit using standard 0.6-μm CMOS technology and constructed an analog memory using the programming circuit and a floating-gate MOSFET. The measurement results indicate that the analog memory attains a programming time of 75 μs, an updating resolution of 11 bit, and a memory setting precision of 6.5 bit. This programming circuit can be used for intelligent information processing hardware such as self-learning VLSI neural networks as well as multilevel flash memory.**

Index Terms—**Analog memory, EEPROM, flash memory, multilevel programming, pulsewidth modulation.**

I. Introduction

ALTHOUGH many VLSI digital storage techniques have been developed, direct analog storage or multilevel digital data storage is more effective in applications such as voice recording [1], [2], neural networks [3], [4], or multivalued logic systems. Multilevel programming in flash memory is also recognized as an effective technique for high-density mass storage [5].

Floating-gate devices can retain analog data for long periods and constitute practical nonvolatile analog memory in the present LSI technology. Therefore, many analog memory circuits and devices using floating-gate structures have been proposed [1], [2], [6]–[11]. In these reports, most programming circuits for conventional floating-gate structures use feedback control in the programming loop consisting of write and read/verify periods [1], [2], [5]. Because the write process usually uses Fowler–Nordheim tunneling, which has exponential characteristics to the applied voltage, write pulses with a constant voltage are used to improve the controllability. In these feedback control methods, since write pulses with a constant width and voltage are used, a tradeoff between programming speed and precision is employed. For example, programming time of a few hundred microseconds is required for achieving 6–8-bit precision [5].

Manuscript received August 22, 2000; revised January 27, 2001. This work was supported in part by Grants-in-aid for the Core Research for Evolutional Science and Technology (CREST) from Japan Science and Technology Corporation (JST).

S. Kinoshita was with Graduate School of Advanced Sciences of Matter, Hiroshima University, Higashi-Hiroshima, 739-8527 Japan. He is now with Rohm Corporation, Ltd., Kyoto, 615-8585 Japan.

T. Morie, M. Nagata, and A. Iwata are with Graduate School of Advanced Sciences of Matter, Hiroshima University, Higashi-Hiroshima, 739-8527, Japan (e-mail: morie@dsl.hiroshima-u.ac.jp).

Publisher Item Identifier S 0018-9200(01)05837-1.

We have already proposed new programming circuits using pulsewidth modulation (PWM) signals for nonvolatile analog memory [12]. We have also demonstrated both fast speed and high precision by means of circuit simulation (HSPICE). The programming circuits can update memory in both incremental and decremental directions, and they also attain high updating resolution, which is important for applications in neural networks [3], [4].

In this paper, we report the measurement results of an analog memory using a floating-gate EEPROM device controlled by one of the programming circuits proposed in [12] that has been designed and fabricated using standard CMOS technology. Section II describes the designed programming circuit. Since this programming circuit generates asynchronous control pulses, it can operate independently. Therefore, it allows faster programming than does the synchronous updating circuit. Section III describes the fabricated chip and its feed-forward control characteristics. Section IV describes the measurement results of the analog memory.

II. A PWM Programming Circuit

Fig. 1 shows an analog memory circuit consisting of the programming circuit, a floating-gate MOSFET, and two high-voltage switches (HVSWs). This programming circuit generated PWM signals based on the voltage difference between the stored data, $V_{\rm OUT}$, and the target, $V_{\rm TAR}$. The floating-gate MOSFET updates its stored value $V_{\rm OUT}$ by means of the PWM signals.

This programming circuit includes a differential amplifier (AMP), a comparator (CMP), a latch circuit (LATCH), and a clock pulse generator (PGEN). As shown in the timing diagram in Fig. 1, the programming operation is asynchronous and autonomous.

The programming operation proceeds as follows. Here, the numbers of the parentheses correspond to those in the timing diagram shown in Fig. 1.

1) Initialize: by clock ϕ_1, switches S_1, S_3 and S_4 are closed. Both inputs of AMP are set at $V_{\rm TAR}$ and node n_3 is set at the threshold voltage of inverter U_1. This operation compensates for the offset of AMP and inverter U_1.
2) Clock pulse ϕ_2 is generated by PGEN: by the termination of clock ϕ_1 (1), node n_8 is set at the *High* level (2) because there is no write pulse at either node n_6 or n_7. Then, capacitor C_R is connected to current source I_3 and is charged up (3). At the same time, clock pulse ϕ_2 appears at node n_{10} (3).
3) Switches S_1 and S_4 are opened by the termination of clock ϕ_1, and switches S_2, S_3 and S_5 are closed by clock

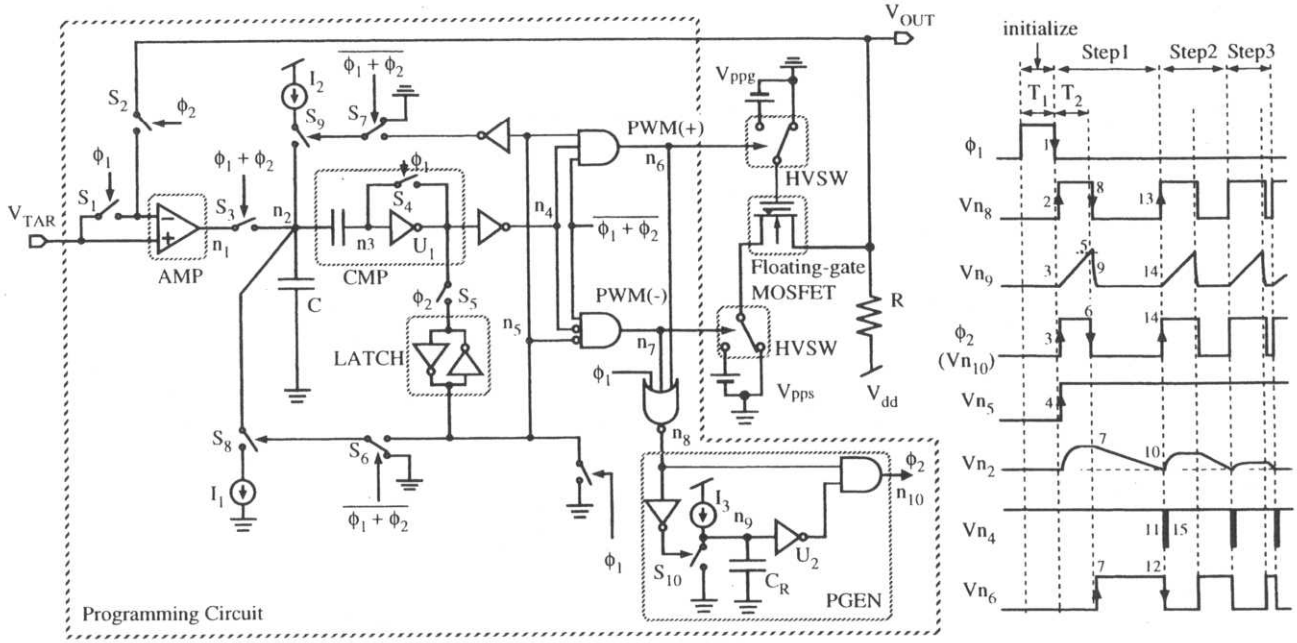

Fig. 1. Memory circuit and its schematic timing diagram when $V_{\mathrm{OUT}} < V_{\mathrm{TAR}}$. The numbers in the timing diagram indicate the order of causality.

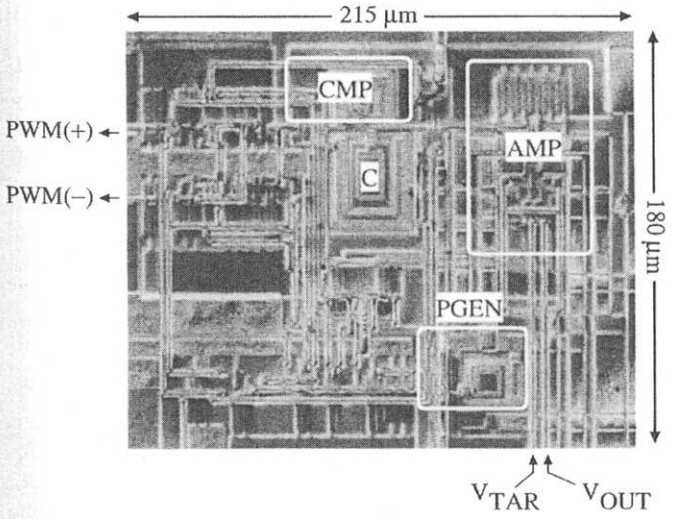

Fig. 2. Chip photograph of the programming circuit.

ϕ_2. Voltages V_{OUT} and V_{TAR} are fed into AMP, and node n_2 is set at the output voltage of AMP.

4) If $V_{\mathrm{OUT}} < V_{\mathrm{TAR}}$, capacitor C is charged up. Comparator CMP detects which larger V_{OUT} or V_{TAR} is, and LATCH holds this information. In this case, node n_4 is set at *High*, and the output of LATCH, namely n_5, also holds the *High* level (4).
5) When the voltage of node n_9 exceeds the threshold voltage of the inverter U_2 (5), pulse ϕ_2 terminates (6). Node n_6 is set at *High* (7). At the same time, switches S_2, S_3 and S_5 are opened and switches S_6 and S_7 are closed. Since node n_5 is *High*, switch S_8 turns on. Capacitor C is connected to current source I_1 and discharged. Thus, the voltage of node n_2 lowers (7). Because n_6 is set at *High*, node n_8 is set at *Low* (8), and capacitor C_R is discharged (9).
6) When the voltage of node n_2 falls, inverter U_1 inverts (10), and node n_4 is set at *Low* (11). Then, node n_6 is also set at *Low* (12). Thus, a PWM write pulse is generated at node n_6. It supplies high-voltage V_{ppg} to the control gate of the floating-gate MOSFET during the period of the pulsewidth, and V_{OUT} rises and approaches V_{TAR}.

On the other hand, if $V_{\mathrm{OUT}} > V_{\mathrm{TAR}}$, capacitor C is initially discharged. Node n_5 is set at *Low* and switch S_9 turns on. Capacitor C is charged up, and then a PWM write pulse appears at node n_7. Voltage V_{OUT} lowers and approaches V_{TAR}. Finally V_{OUT} becomes equal to V_{TAR} by repeating this operation. The offset compensation operation triggered by clock ϕ_1 is performed only once at the first step of the entire programming operation.

The programming speed and resolution depend on the gain of AMP as well as the characteristics of current sources I_1 and I_2. Therefore, these characteristics should be carefully adjusted. On the other hand, the saturation characteristic of AMP reduces the programming speed when V_{OUT} is far from V_{TAR} because the width of the PWM updating signals is limited. However, this nonlinear characteristic is useful because the gain can be adjusted arbitrarily irrespective of the limited output range of AMP.

III. Fabricated Chip and Its Performance

A. Fabricated Chip

We fabricated the programming circuit by using a 0.6-μm double-poly triple-metal CMOS process. Fig. 2 shows the micro-photograph of the programming circuit. The circuit size is $215 \times 180\ \mu\mathrm{m}^2$.

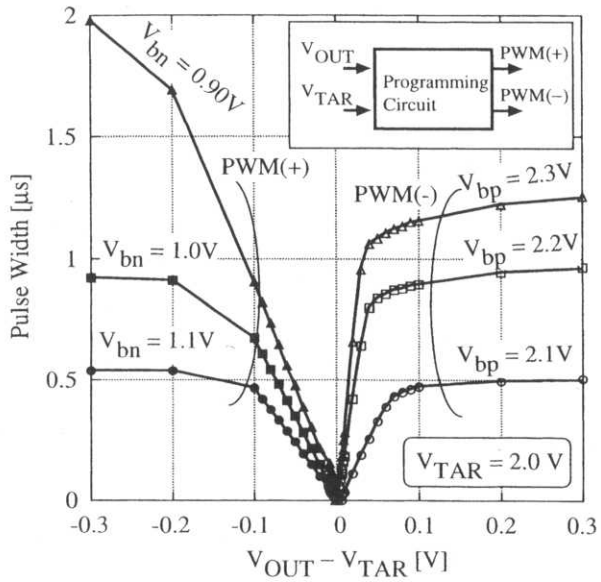

Fig. 3. Measurement results of the programming circuit in feedforward operation, which shows the relationship between $(V_{\rm OUT} - V_{\rm TAR})$ and the pulsewidth of PWM output signals. Voltages $V_{\rm bn}$ and $V_{\rm bp}$ are the bias voltages of n- and p-MOSFETs used as current sources I_1 and I_2 shown in Fig. 1, respectively, which are the parameters related to the pulsewidth conversion ratio. PWM(+) indicates PWM pulses increasing $V_{\rm OUT}$, and PWM(−) indicates those decreasing $V_{\rm OUT}$.

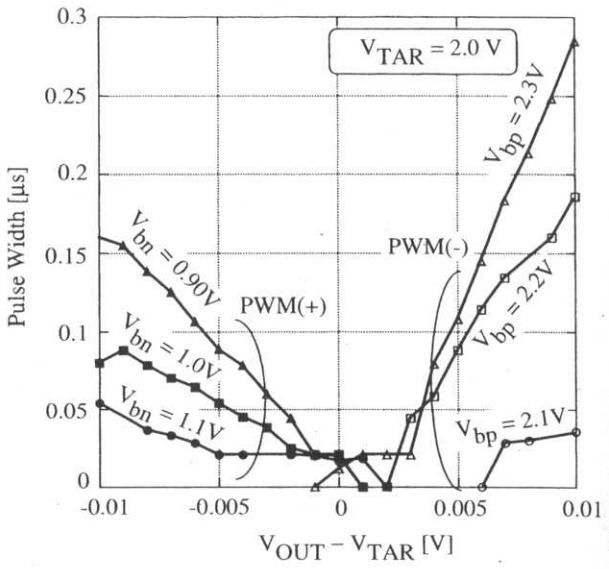

Fig. 4. Magnification of the region near $V_{\rm OUT} - V_{\rm TAR} = 0$ in Fig. 3.

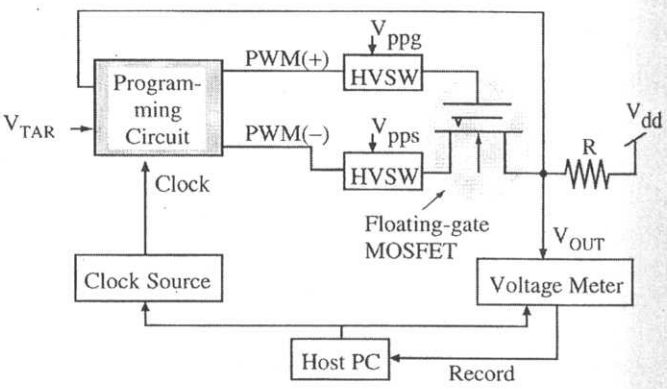

Fig. 5. Experimental setup for analog memory operation using the fabricated programming circuit and a floating-gate MOSFET.

B. Feedforward Control Characteristics of the Fabricated Circuit

We measured the feedforward characteristics of the fabricated programming circuit. We fed voltages $V_{\rm OUT}$ and $V_{\rm TAR}$ into the programming circuit as the input signals, and observed the PWM output signals. The supply voltage was 3.3 V.

Fig. 3 shows the measurement results of the programming circuit, which demonstrates the relationship between $(V_{\rm OUT} - V_{\rm TAR})$ and the pulsewidth of the PWM output signals. The saturation characteristics shown in Fig. 3 reflect the characteristic of AMP. When $(V_{\rm OUT} - V_{\rm TAR})$ is large, the pulsewidth is almost maximum and constant, and therefore the update is maximum. When the difference is small, the pulsewidth is proportional to the difference. Therefore, $V_{\rm OUT}$ precisely approaches $V_{\rm TAR}$.

Fig. 4 magnifies the region near $V_{\rm OUT} - V_{\rm TAR} = 0$ in Fig. 3. The minimum pulsewidth is about 20 ns, and the minimum value of $|V_{\rm OUT} - V_{\rm TAR}|$ is about 1 mV. If the full scale of $V_{\rm OUT}$ is assumed to be 1 V, this indicates 10-bit resolution.

IV. Analog Memory Performance of a Floating-Gate MOSFET Controlled by the Fabricated Programming Circuit

A. Experimental Setup

We measured the performance of analog memory using a floating-gate MOSFET controlled by our fabricated programming circuit. The floating-gate MOSFET used was fabricated by using a conventional EEPROM process, and its gate length and width were 0.75 μm and 2.2 μm, respectively.

Fig. 5 shows the experimental setup. Each of high voltage switches (HVSWs) consists of two resistive-load NMOS inverters using commercially available MOSFETs (2SK2731) and load resistance of 100 Ω. PWM pulses PWM(+) and PWM(−) are fed into the HVSWs, and converted into high-voltage PWM signals. The high voltage $V_{\rm ppg}$ or $V_{\rm pps}$ is supplied to the floating-gate MOSFET during the period of the pulsewidth. Thus, $V_{\rm OUT}$ is updated and approaches $V_{\rm TAR}$. In order to achieve fast and high-precision programming, we performed asymmetric programming by setting $V_{\rm ppg} \neq V_{\rm pps}$, i.e., the update rate was set large when $V_{\rm OUT} > V_{\rm TAR}$, and it was set small when $V_{\rm OUT} < V_{\rm TAR}$ [12]. We recorded the number of pulses applied to the floating-gate MOSFET, $N_{\rm pulse}$, and the memorized voltage $V_{\rm OUT}$.

B. Experimental Results and Discussion

Fig. 6 shows the programming performance: the relationship between $N_{\rm pulse}$ and $V_{\rm OUT}$. The solid line shows a programming result of our system shown in Fig. 5. Since the asymmetric programming method was used, this result shows the worst case from the viewpoint of programming speed. At the beginning of programming ($V_{\rm OUT} = 1.6$ V), the pulsewidth of PWM write signals was 1.8 μs. On the other hand, when $V_{\rm OUT}$ approached to $V_{\rm TAR}$ ($V_{\rm OUT} = 2.6$ V), the pulsewidth was 200 ns.

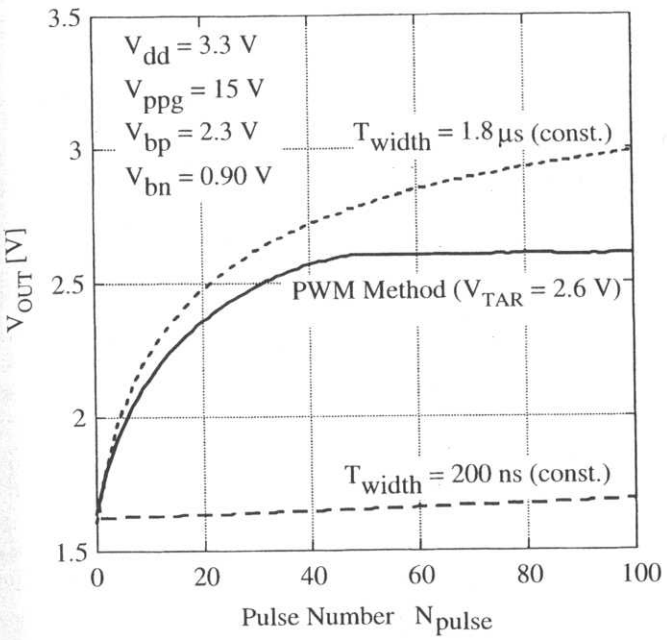

Fig. 6. Programming performance: relationship between the number of write pulses $N_{\rm pulse}$ and $V_{\rm OUT}$. The solid line shows a programming result of our system. The two broken lines show results with constant-width pulses having the initial pulsewidth (1.8 μs) and the final one (200 ns), respectively.

In order to clarify the effectiveness of our variable-pulsewidth programming method, we also performed the conventional programming method using constant-width pulses, where the write pulsewidths $T_{\rm width}$ are 1.8 μs and 200 ns. The results are also shown in Fig. 6.

Our PWM method required 50 pulses in order that $V_{\rm OUT}$ converges to $V_{\rm TAR}$, which indicates a programming time of 75 μs. When we used constant pulses whose width ($T_{\rm width}$) was 1.8 μs, $V_{\rm OUT}$ arrived at $V_{\rm TAR}$ by 30 pulses. This is faster than the PWM method, but a large update value of 14 mV leads to lower resolution around 6 bit when the full scale is assumed to be 1 V. In contrast, the PWM method can update with a resolution of about 0.5 mV near $V_{\rm TAR}$, which corresponds to 11-bit resolution.

On the other hand, if we want to obtain an update resolution of 0.5 mV using constant pulses, their width must be 200 ns. In this case, 3500 pulses are required to reach $V_{\rm TAR}$. This means that the programming time is more than 700 μs. Thus, the PWM method can achieve analog programming ten times faster than the method using constant-width pulses.

It is noted that the relation between total pulsewidth ($T_{\rm width} \times N_{\rm pulse}$) and $V_{\rm out}$ is different for the case of $T_{\rm width} = 200$ ns and that of $T_{\rm width} = 1.8$ μs. This is because the output pulses of the HVSWs have gradual rise characteristics, which is usually required for improving the reliability of floating-gate devices in the write/erase operation.

Fig. 7 shows the setting error of $V_{\rm OUT}$ as a function of $V_{\rm TAR}$. Because of asymmetric updating, $V_{\rm OUT}$ always approach $V_{\rm TAR}$ under the condition that $V_{\rm OUT} < V_{\rm TAR}$. If the controllable range of $V_{\rm OUT}$ is assumed to be from 1.6 V to 2.6 V, the maximum error is 11 mV. This achieves 6.5-bit precision.

Here, let us discuss the temperature dependence of the performance. If the readout temperature differs from the programming temperature, an error can occur because of the temperature dependence of the floating-gate MOSFET characteristics. This error, however, can be minimized by using a MOSFET that has the same temperature dependence as the floating-gate MOSFET instead of the resistor R. This is because $V_{\rm OUT}$ is determined by the ratio between the ON-resistance of the floating-gate MOSFET and the resistance R.

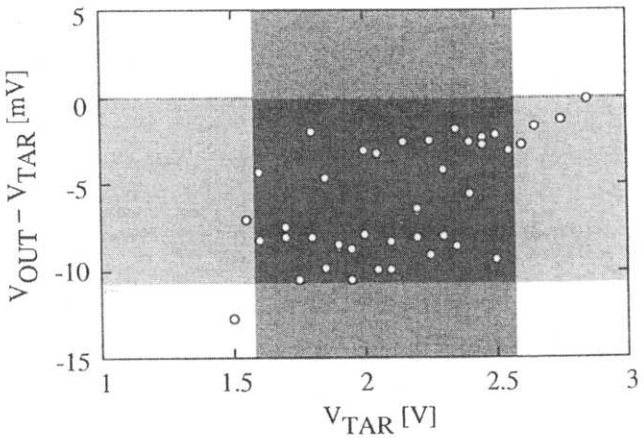

Fig. 7. Setting error of $V_{\rm OUT}$ as a function of $V_{\rm TAR}$. $V_{\rm OUT}$ is always lower than $V_{\rm TAR}$ because of asymmetric updating.

The retention time of the floating-gate memory device also depends on the ambient temperature. For analog memory operation, this effect was evaluated using real devices. It has been reported that the retention accuracy was about 6 bit for 15 years at more than 125 °C [4]. Although this result was obtained using a different device from that used in the experiment described here, it supports our assertion that the analog memory described here can maintain the above setting precision over a wide range of ambient temperatures.

V. Conclusion

We measured the performance of analog memory consisting of our PWM programming circuit fabricated using standard CMOS technology and a floating-gate MOSFET fabricated using EEPROM technology. The measurement results showed 6.5-bit setting precision and 11-bit updating resolution when the voltage setting range was assumed to be 1 V. Its programming speed was 75 μs in the worst case using the asymmetric updating operation. A comparison with the experimental results using constant-width write pulses confirmed that our PWM programming method has 30 times finer updating resolution or 10 times faster programming speed than conventional methods.

Acknowledgment

The authors wish to thank Prof. M. Hirose for his support and encouragement. The floating-gate devices were specially offered by Rohm Co., Ltd. The PWM programming chip has been fabricated in the chip fabrication program of VLSI Design and Education Center (VDEC), University of Tokyo, with the collaboration by Rohm Corporation and Toppan Printing Corporation.

REFERENCES

[1] T. Blyth, S. Khan, and R. Simko, "A nonvolatile analog storage device using EEPROM technology," in *IEEE Int. Solid-State Circuits Conf. Dig.*, 1991, pp. 192–193.

[2] H. V. Tran, T. Blyth, D. Sowards, L. Engh, B. S. Nataraj, T. Dunne, H. Wang, V. Sarin, T. Lam, H. Nazarian, and G. Hu, "A 2.5 V 256-level nonvolatile analog storage device using EEPROM technology," in *IEEE Int. Solid-State Circuits Conf. Dig.*, 1996, pp. 270–271.

[3] T. Morie and Y. Amemiya, "An all-analog expandable neural network LSI with on-chip backpropagation learning," *IEEE J. Solid-State Circuits*, vol. 29, pp. 1086–1093, Sept. 1994.

[4] T. Morie, O. Fujita, and K. Uchimura, "Self-learning analog neural network LSI with high-resolution nonvolatile analog memory and a partially-serial weight-update architecture," *IEICE Trans. Electron.*, vol. E80-C, no. 7, pp. 990–995, 1997.

[5] P. L. Rolandi, R. Canegallo, E. Chioffi, D. Gerna, G. Guaitini, C. Issartel, F. Lhermet, M. Pasotti, and A. Kramer, "1 M-cell 6 b/cell analog flash memory for digital storage," in *IEEE Int. Solid-State Circuits Conf. Dig.*, 1998, pp. 334–335.

[6] O. Fujita, Y. Amemiya, and A. Iwata, "Characteristics of floating gate device as analogue memory for neural networks," *Electron. Lett.*, vol. 27, pp. 924–926, 1991.

[7] O. Fujita and Y. Amemiya, "A floating-gate analog memory device for neural networks," *IEEE Trans. Electron Devices*, vol. 40, pp. 2029–2035, Nov. 1993.

[8] M. Lanzoni, L. Briozzo, and B. Riccò, "A novel approach to controlled programming of tunnel-based floating-gate MOSFET's," *IEEE J. Solid-State Circuits*, vol. 29, pp. 147–150, Feb. 1994.

[9] K. Nakajima, S. Sato, T. Kitaura, J. Murota, and Y. Sawada, "Hardware implementation of new analog memory for neural networks," *IEICE Trans. Electron.*, vol. E78-C, no. 1, pp. 101–105, 1995.

[10] T. Shibata, T. Nakai, N. M. Yu, Y. Yamashita, M. Konda, and T. Ohmi, "Advances in neuron-MOS application," in *IEEE Int. Solid-State Circuits Conf. Dig.*, 1996, pp. 304–305.

[11] K. Kim and K. Lee, "An 8 b resolution 360 μs write time nonvolatile analog memory based on differentially balanced constant-tunneling-current scheme (DBCS)," in *IEEE Int. Solid-State Circuits Conf. Dig.*, 1998, pp. 336–337.

[12] S. Kinoshita, T. Morie, M. Nagata, and A. Iwata, "New nonvolatile analog memory circuits using PWM methods," *IEICE Trans. Electron.*, vol. E82-C, no. 9, pp. 1655–1661, 1999.