Extended Abstracts of the 1994 International Conference on Solid State Devices and Materials, Yokohama, 1994, pp. 346-348

A Neuron-MOS Neural Network Using Low-Power Self-Learning-Compatible Synapse Cells

Tadashi SHIBATA, Hideo KOSAKA, Hiroshi ISHII, and Tadahiro OHMI

Department of Electronic Engineering, Tohoku University
Aza-Aoba, Aramaki, Aobaku, Sendai 980 JAPAN

A self-learning neural network hardware has been developed using Neuron MOS Transistor (νMOS) as a key circuit element, which is a functional device simulating the actions of biological neurons at a single transistor level [1]. Synapse cells are formed by merging an EEPROM memory cell into a new-concept νMOS differential-source-follower circuitry. As a result, synaptic connections free from standby power dissipation and featuring excellent weight-updating characteristics have been established. The operation of the synapse cells and νMOS neural networks has been verified using test circuits fabricated by a double-polysilicon CMOS process. An interesting feature of the synapse cell, the acceleration effect in learning, is also presented.

1. INTRODUCTION

Neural networks are now drawing a considerable attention as a new paradigm of information processing due to their self-adaptive problem-solving capabilities. The essence of the adaptive learning of a neural network is the modification of the synaptic weights to establish desirable response of the system in solving problems. Therefore the development of an electronic version of the synaptic connection compatible to learning algorithms is most essential in developing the self-learning hardware of neural networks.

The purpose of this study is to develop neural networks using Neuron MOS Transistor (νMOS)[1] as a key circuit element. In order to establish an on-chip self-learning function on νMOS neural networks, a new synapse memory cell has been developed [2] and employed in this work. The circuit implementation of a hardware-oriented learning algorithm (Hardware Backpropagation) [3] and the verification of its powerful capability in solving problems [4] are presented in separate articles.

2. NEURAL NETWORK CONFIGURATION

A single neuron module is schematically shown in Fig. 1 where a neuron is composed of a complementary νMOS inverter having all identical coupling capacitors and a regular CMOS inverter. $V_1 \sim V_n$ are the outputs of previous-layer neurons, being multiplied by the respective weights at synapses, and then, transferred to the neuron inputs. Since the neuron has hard-limiter characteristics, V_i takes either a binary 1 or a binary 0. Then the function of the synapse is just transferring the

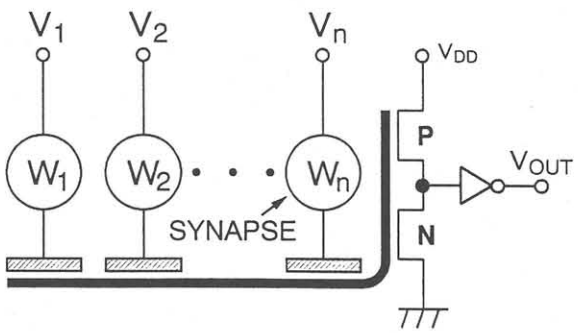


Fig. 1. Schematic of a neuron module composed of a complementary νMOS inverter and synapse cells.

weight W_i to a neuron input when $V_i = 1$. The net = $\Sigma W_i V_i$ is automatically calculated by charge redistribution on the floating gate and then squashed into a 0 or a 1.

The synapse cell is shown in Fig. 2, where Tr. 1 is always ON in the forward network operation. The charge on the synapse floating gate Q_F^S representing the synaptic weight is non-destructively read out by source follower actions of both N- and P-νMOS' which share the common floating gate. These source followers are merged into CMOS inverters to cut off the dc current paths, thus achieving the standby-power free feature of the cell. The output voltages of both N- and P-νMOS source followers are transferred to the common neuron floating gate (dendrite) via capacitance coupling, thus being averaged on the dendrite.

The measured output wave forms of N- and P-νMOS source followers (V^+ and V^-, respectively, on the figure) are shown in Fig. 3. The effective synapse output

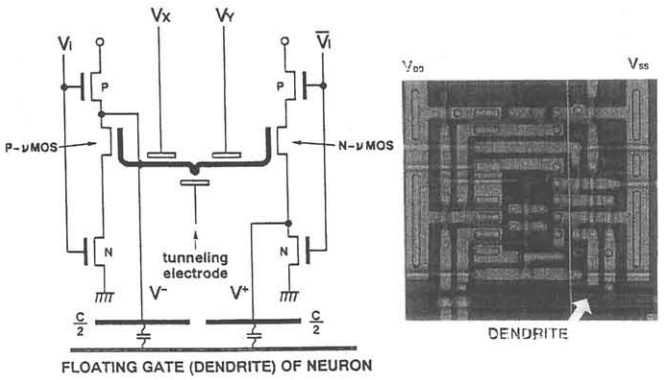


Fig. 2. Complementary νMOS differntial source-follower circuit for synapse cell.

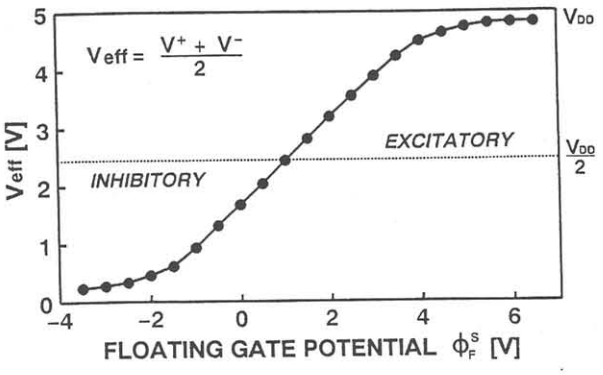


Fig. 4. Measured synapse output as a function of the synapse floating-gate potential.

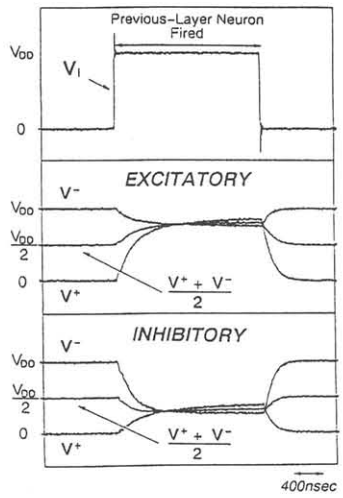
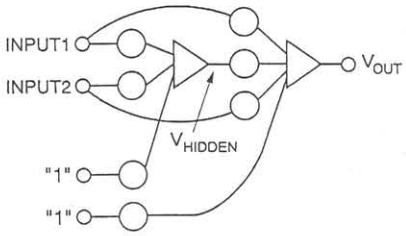


Fig. 3. Measured output wave forms of N-νMOS (V+) and P-νMOS (V-) source followers in the synapse cell.



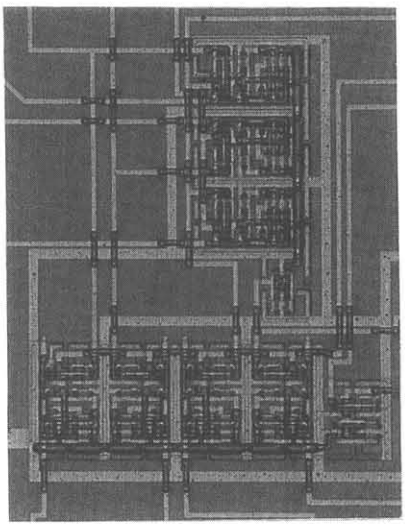


Fig. 5. Network configuration and photomicrograph of test νMOS neural network.

$V_{eff} = (V^+ + V^-)/2$ vs. synapse floating-gate potential ϕ_F^S is shown by experimental data in Fig. 4. The cell can represent either an excitatory synapse (positive weight) or an inhibitory synapse (negative weight) depending on whether V_{eff} is larger or smaller than $V_{DD}/2$, the neutral output of a synapse, respectively. Since $\phi_F^S = Q_F^S/C_{TOT}^S$ (C_{TOT}^S: the total capacitance of the synapse floating gate), the weight value can be altered by electron injection or extraction through the tunnel oxide by giving programming pulses to both V_X and V_Y. Since the programming occurs only at the crossing points of V_X and V_Y lines running over the synapse cell array, a Hebbean-like learning algorithm can be easily implemented. Such a selective cell programming technique was first introduced in a dual-control-gate EEPROM cell [5].

3. NEURAL NETWORK OPERATION

In order to verify the operation of νMOS neural networks, a simple test circuit was designed and fabricated by a double polysilicon CMOS process. Fig. 5 shows the network configuration and a photomicrograph of the circuit. In this circuit, the floating gates of synapses are directly connected to external electrodes and their potentials were set by external voltage sources. The weight values were determined by learning simulator based on the Hardware Backpropagation (HBP) algorithm [3,6]. Fig. 6 shows the measured circuit responses when it learned the XOR or OR functions, demonstrating the correct operation of the νMOS neural network.

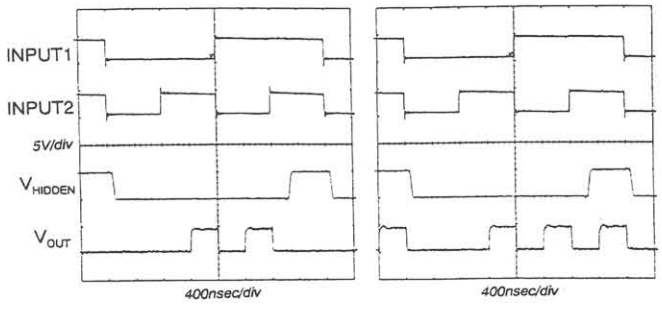


Fig. 6. Measured output characteristics of the test neural network which learned XOR and OR Boolean functions.

4. WEIGHT UPDATING CHARACTERISTICS

EEPROM cell writing by applying constant voltage pulses exhibits a strong nonlinear dependence on the number of pulses due to the non-linear J-V characteristics of a Fouler-Nordheim tunneling current. This is one of the most critical issues of using EEPROM technology for synapses [7].

The synapse cell shown in Fig. 2 has solved the problem very beautifully by just adding an NMOS transistor (Tr. 1) to our original six-transistor-version synapse cell [6]. When Tr. 1 is on, the output of the N-νMOS source follower is fed back to the tunneling electrode and reset the voltage across the tunneling oxide always at a constant value of V_{TN} (the threshold of N-νMOS) indifferent to the amount of charge stored on the floating gate. This assures the constant charge injection (or extraction) under a constant programming pulse. Fig. 7 compares the weight updating characteristics of a conventional cell (no feed back) and the new cell, where the excellent linearity is evident for the new cell. Such a feature is quite essential for hardware learning of neural networks.

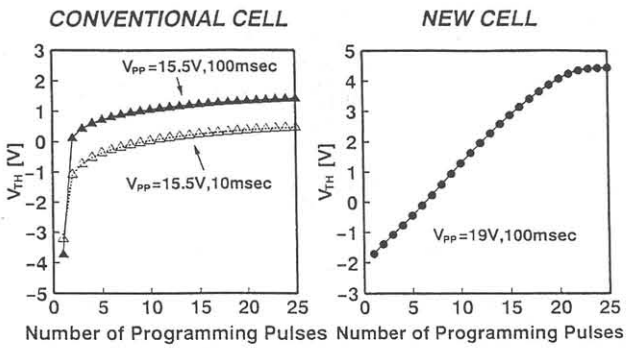


Fig. 7. Comparison of weight-updating characteristics of conventional cell (without source follower feedback) and new synapse cell (with feedback).

Further advantage of the new cell is demonstrated in Fig. 8, where the measured increment in the cell threshold voltage ΔV_{TH} (equal to $-\Delta Q_F^S/C_{TOT}^S$) is plotted against V_{TH}. The data indicates that the larger the total number of electrons in the floating gate is, that the more electrons are injected per pulse. This phenomenon is quite unnatural and has been realized for the first time in our new cell by optimizing the intra-cell coupling capacitance ratios. This provides a means to implement the acceleration effect in weight updating, which is advantageous in enhancing the learning performance of the circuit as shown by the results of learning simulation in Fig. 9.

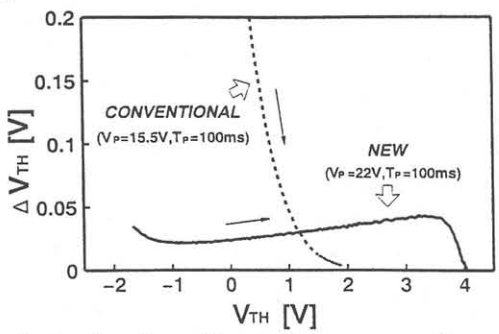


Fig. 8. Acceleration effect in the synapse cell.

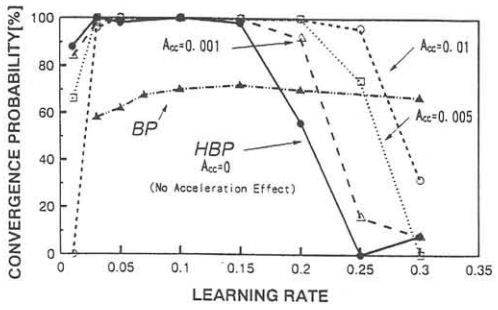


Fig. 9. Learning performance of three-input Exclusive OR by HBP for varying acceleration coefficients. Ordinate represents the convergence probability of learning for 50 different initial random-weight patterns.

5. CONCLUSIONS

The construction of a self-learning neural network using νMOS technology has been discussed and the basic design concepts have been verified by experiments. The excellent linearity in weight updating characteristics or even the acceleration effects provided by the new synapse circuits are able to be used very favorably in enhancing the learning performance of the hardware. νMOS neural networks operates purely in the voltage mode, making the approach much superior to conventional current-mode approach in terms of low-power dissipation, which is the most important requirement of ultralarge scale integrated systems.

Acknowledgment

This work was partially supported by the Ministry of Education, Science and Culture of Japan under the Grant-in-Aid for Developmental Scientific Research (No. 05505005). Major part of this work was carried out in the Superclean Room of the Laboratory for Microelectronics, Research Institute of Electrical Communication, Tohoku University.

References

[1] T. Shibata and T. Ohmi, IEEE Trans. Electron Devices, **39**, 1444 (1992).

[2] H. Kosaka, T. Shibata, H. Ishii, and T. Ohmi, IEDM Tech. Dig., 623 (1993).

[3] H. Ishii et al., this conference.

[4] S. Kondo et al., this conference.

[5] K. Hieda et al., IEEE Trans. Electron Devices, **ED-32**, 1776 (1985).

[6] H. Ishii, T. Shibata, H. Kosaka, and T. Ohmi, IEDM Tech. Dig,. 435 (1992).

[7] O. Fujita, Y. Amemiya, and A. Iwata, Electronics Lett., **27**, 924 (1991).