Extended Abstracts of the 1993 International Conference on Solid State Devices and Materials, Makuhari, 1993, pp. 449-451

Switched Diffusion Analog Memory for Neural Networks

T.KITAURA, S.SATO, K.NAKAJIMA, J.MUROTA, AND Y.SAWADA

Laboratory for Microelectronics,
Research Institute of Electrical Communication, Tohoku University,
2-1-1 Katahira, Aoba-ku, Sendai 980, JAPAN

We have fabricated a new analog memory with a floating-gate as a key component to store synaptic weights for integrated artificial neural networks. The new analog memory comprises a tunnel junction (poly-Si/poly-Si oxide/poly-Si sandwich structure) , a thin-film transistor, and a floating-gate MOSFET. The diffusion of the charges injected through the tunnel junction are controlled by switching operation of the thin-film transistor, and we refer to the new analog memory as 'Switched Diffusion Analog Memory (SDAM)'. The obtained characteristics of an SDAM shows a fast switching speed and an improved linearity between the potential of the floating-gate and the number of pulse inputs. SDAM can be used in a neural network in which write/erase and read operation performs simultaneously.

1 INTRODUCTION

Artificial neural networks provide a specific approach to distributed parallel processing based on actual neurophysiological models. Recently there has been active studies for VLSI implementation of neural networks[1] , which require huge number of memory devices storing synaptic weights to accomplish a learning process. A key technology for hardware integration is how to implement a large number of such memory devices into a small area. In this research, an EEPROM has a large potential as an analog memory for neural networks, because an analog memory is superior in area occupancy to a digital memory. But an existing EEPROM cell has following weak points which require a complex circuitry for learning; the floating-gate potential varies in inverse proportion to logarithm of number of input pulses, and write/erase operation requires a separate time period from the network activity. In the present paper, we report fabrication of a new analog memory and improved characteristics measured. The new analog memory comprises a tunnel junction, a thin-film transistor(TFT), and a MOSFET which were designed in nMOS technology with 4μm design rules.

2 DEVICE DESIGN

An improved structure of a MOSFET with a floating-gate has been proposed by Fujita[2]. The floating-gate is divided into two parts which are connected with a high resistance. But the high resistance reduces the speed of the write operation into the floating-gate. To improve the operation speed, we have designed a new analog memory with a floating-gate which is divided into three parts, two capacitors and a TFT as shown in Fig.1.

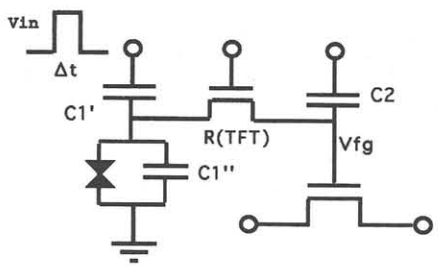

Figure 1: Equivalent circuit of the new analog memory.

Figure 1 shows an equivalent circuit of the new analog memory with a connection TFT. One of the three parts of the floating-gate is a small capacitance with a tunnel junction for charge injection, the second part is a connection TFT for charge transfer, and the third part is a large capacitance for charge storage.

Device operations of the new analog memory are the followings. When the TFT is off, the read-out operation using the MOSFET and the charge injection can be done simultaneously. The amount of the injected charges per pulse is quantized by the small capacitance of the injection part. The charges can diffuse rapidly to the whole floating-gate, when the

TFT is on. These functions of the analog memory are desirable to hold synaptic weights especially for pulse-output neural networks, one of which we have already reported[3].

Assuming that the input pulse width Δt is long enough to charge up the small capacitor at the injection part through a tunnel junction, the potential of the floating-gate V_{fg} is given by,

$$V_{fg} = V'_{in} - (V'_{in} - V_0)\exp^{-(\frac{\Delta t}{C_2 R} + \ln\frac{C_1+C_2}{C_2})n} \quad (1)$$

where V'_{in} is the saturation voltage which is determined by the characteristic of the tunnel junction, $C_1 = C'_1 + C''_1$ is the total capacitance of the charge injection part as shown in Fig.1, R is the on-resistance value of the TFT, and V_0 is an initial value of V_{fg} . It can be seen from Eq.(1) that the smaller the ratio C_1 to C_2 is, the better the linearity between V_{fg} and n is, although V_{fg} changes exponentially with the number of pulse inputs n. The charges injected into the small capacitor are diffused by switch-on operation of the TFT, and therefore the new analog memory is referred to as 'Switched Diffusion Analog Memory (SDAM)' in this paper.

3 STRUCTURE AND IMPLEMENTATION

Figure 2 shows the structure of the fabricated analog memory SDAM. We used a thermal oxide film of poly-Si as the tunneling barrier for the charge injection part. Because the oxide film for tunneling is generally thicker than that of single-crystal Si, which results in a smaller capacitance. A small ratio of C_1 to C_2 improves the linearity between the floating-gate voltage and the input pulse number. All components of the analog memory(the TFT, the tunnel junction, the capacitors, and the MOSFET) were formed by the depositions of two poly-Si films. The channel of the TFT was formed by the undoped part of the first poly-Si layer. The ion implantation for the TFT and the MOSFET were carried out simultaneously in the present process. Figure 3 shows a photograph of the fabricated analog memory SDAM.

4 EXPERIMENTALS

4.1 Process

The junction size is 4μm by 4μm, and the thickness of the tunnel barrier is about 200Å. We made the poly-Si layers for the tunnel junctions under three kinds of different conditions where the gas sources are Si_2H_6 at the deposition temperature of 550℃, SiH_4 at 600℃, and SiH_4 at 650℃. The arsenic ion implantation for the first poly-Si layer and source/drain of MOSFET has been performed at

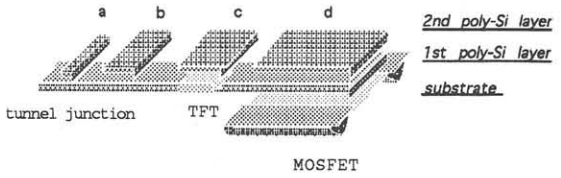

Figure 2: Structure of the fabricated new analog memory. a: tunnel gate, b: bias gate, c: TFT gate, and d: control gate.

100keV with a dose of 1x10^{15}cm^{-2}, simultaneously. The poly-Si oxide layer has been formed in a dry oxygen atmosphere at 800℃.

4.2 Tunnel characteristics of poly-Si/poly-Si oxide/poly-Si

Figure 4 shows I-V characteristics of the tunnel junctions made of poly-Si. Breakdown voltages of the junctions are around 10V. The junction made from Si_2H_6 has the highest tunneling voltage, because the deposition film was amorphous and it had a smooth surface. The emission seems to be of the Fowler-Nordheim tunneling type as shown in Fig.5.

4.3 SDAM characteristics

The stored charges in an SDAM change the drain current of the component MOSFET of the SDAM. Figure 6 shows the results measured for the drain current as a function of the number of input pulses (injection/extraction) for four values of the pulse widths. The input pulse amplitude, the switching voltage for the TFT, and the source-drain and the control gate voltages of the component MOSFET

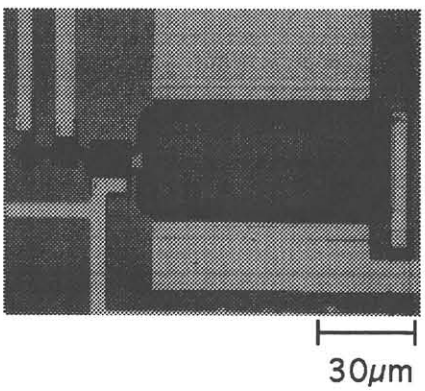

Figure 3: Photograph of the fabricated new analog memory.

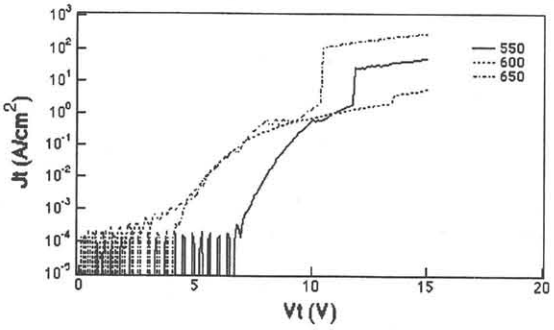

Figure 4: I-V characteristics of the tunnel junctions(the oxide thicknesses are 200Å) which have been fabricated in the deposition of SiH_4 at 600℃ and 650℃, Si_2H_6 at 550℃.

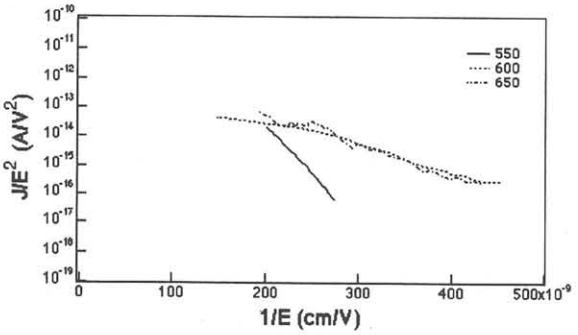

Figure 5: Fowler-Nordheim plots of the tunneling currents shown in Fig.4.

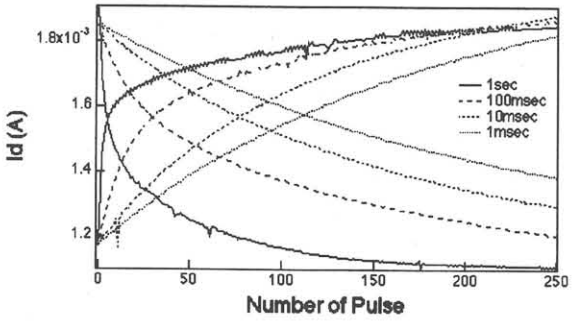

Figure 6: Experimental result of the drain current as a function of the number of input pulses (injection/extraction) for four values of the pulse widths.

are 16V, 8V, 5V and 8V, respectively. The result shown in Fig.6 shows that the nonlinearity in the relation between the drain current and the number of pulse inputs increases with increasing pulse width Δt, for Δt longer than 1msec. The experimental result agrees with the numerical simulation, which shows that the amount of injected charges are determined only by the capacitance of the injection part independently of the pulse width Δt, for Δt less than 1msec.

5 CONCLUSION

We fabricated a new type of analog memory with a floating-gate which is used as a key component to store synaptic weights for integrated artificial neural networks. The analog memory comprises a tunnel junction, a connection TFT, and a MOSFET. A small capacitance realized by use of poly-Si oxide film as the tunneling barrier together with the switching of TFT improved linearity. The experimental result agrees with the numerical simulation which shows an improvement of the linearity. The present analog memory realized a fast switching speed and the simultaneous function of the charge injection and the read-out operation.

ACKNOWLEDGEMENTS

This research has been mainly carried out at the Superclean Room of the Laboratory for Microelectronics, Research Institute of Electrical Communication, Tohoku University.

References

[1] For examples, C.Mead, Analog VLSI and neural systems, Addison Wesley, (1989)

[2] O.Fujita, Y.Amemiya, and A.Iwata , *Electronics letters*, vol.27, no.11, p.924 (1991)

[3] S.Sato, M.Yumine, T.Yama, J.Murota, K.Nakajima, and Y.Sawada , *Proc. IJCNN-92*, vol.1, p.172, (1992)