Extended Abstracts of the 2015 International Conference on Solid State Devices and Materials, Sapporo, 2015, pp844-845

Pure CMOS One-Time Programmable Memories using a combination of Gate-Ox Anti-fuse and Poly-Si fuse

Mari Matsumoto, Kosuke Tatsumura, Koichiro Zaitsu and Shinichi Yasuda

Advanced LSI Technology, Corporate R&D Center, Toshiba Corporation
1, Komukai-Toshiba-cho, Saiwai-ku, Kawasaki 212-8582, Japan
Phone: +81-44-549-2188 E-mail: mari.matsumoto@toshiba.co.jp

Abstract

One-time programmable (OTP) memories using CMOS transistors are demonstrated. With the characteristics of gate-oxide anti-fuse and poly-Si fuse, a permanent conductive path is formed between source and drain. Since the programming terminal is separated from the gate, this method realizes high programming efficiency. The cell can be realized entirely by CMOS logic process without extra masks or processing steps.

1. Introduction

Recently, a decrease in LSI manufacturing cost has been required. The conventional embedded non-volatile memories need additional masks or process steps, which increases the manufacturing cost. For this reason, non-volatile memories using standard CMOS transistors have been studied [1-4]. In particular, OTP memory using gate oxide anti-fuse has been reported [1-3]. However, the conventional anti-fuses combine a selection device with an oxide layer, which makes the peripheral circuits complicated. For low-cost LSI, we examined the OTP memory that has a conductive path between the S/D brought about by local breakdown in PN junction [4]. However, large negative voltage is applied to the gate for programing, which requires deep N-well process.

In this work, we propose a novel OTP memory device. By combining the features of gate-oxide anti-fuse and gate poly-Si line fuse, the OTP memory forms a permanent conductive path between source and drain (S/D) in nMOSFET transistor. Programming is controlled by positive gate voltage. Using this memory with 50 nm gate length, we can program at a lower voltage 5 V. In addition, since this OTP memory is fabricated without an additional mask, an LSI in which OTP memories are embedded can be realized low cost.

2. Programming method

The programming method using nMOSFET is illustrated in Fig.1. Figure 2 shows the change in the drain current (I_D) and the gate current (I_G) brought about by applying program voltage V_{PGM} to the gate. V_{PGM} is applied to the gate with S/D both set at 0V via the gate poly-Si line. This causes a breakdown of gate oxide film to electrically connect the gate and the source (G/S), and the gate and the drain (G/D). I_S and I_D successfully change simultaneously (Fig.3 (a)). This means that V_{PGM} makes the conductive paths G/S and G/D. V_{PGM} is then applied to the gate to cause an overcurrent to flow through the gate poly-Si line, the gate electrode, the gate oxide, and the source and drain region. V_{PGM} breaks the gate poly-Si line to prevent the current from flowing to the gate electrode. As a result, a conductive path is formed between the S/D, and programming is controlled by gate voltage. After gate poly-Si line breakdown, I_D and I_S increase linearly with V_D (Fig.3 (b)). Figure 4 shows a scanning electron microscope (SEM) image of an OTP memory after the gate line disconnection. There is a void at the narrow part of the gate poly-Si line. This figure shows that an overcurrent is brought about electro-migration at the poly-Si line.

3. Experiments and analysis

Next, we show fundamental experiments and analyses using the transistors with gate width of 1 um. The gate poly-Si line breaks at the voltage smaller than gate-ox breakdown voltage (Fig.5 (a)). The current decreases for narrower poly-Si line width (Fig.5 (b)). Figure 6 shows programming time T_{PGM1} and T_{PGM2} for various compliance currents (I_{comp}). Here, T_{PGM1} is the time when gate oxide is broken down, and T_{PGM2} is the time when gate poly-Si line is broken. Regardless of I_{comp}, T_{PGM1} decrease exponentially for larger V_{PGM}. On the other hand, T_{PGM2} depends on I_{comp}, when I_{comp} is less than 2 mA. This is because the resistance of conductive path in gate oxide is determined by I_{comp}. The resistance distribution between the S/D after gate line breakdown is shown in Fig.7. Here, R_{SD} (R_{DS}) is the resistance that the voltage applied to source (drain). Regardless of the direction in which the voltage is applied, these distributions are similar. Figure 8 shows that R_{SD} (or R_{DS}) is lower than the resistance R_G after gate oxide breakdown. After gate oxide breakdown, the conductive paths in gate oxide film become wide by applying V_{PGM}. A low resistance leads to a reduction of a delay time in a circuit.

Figure 9 shows an example of an OTP memory array for program and read operation. Thanks to the separation of write terminal (gate) and read terminal (S/D), NAND type structure is feasible. To program the selected cell, the S/D are biased at 0V, and the gate is biased at V_{PGM}. For unselected cells, the S/D are biased at V_{INH} (>0). Even if V_{PGM} is applied to the gate, the cell could not program with V_{INH} at S/D. Figure 10 shows the programming time with the bias condition of a selected cell and an unselected cell. The T_{PGM2} of the selected cell is 10^5 times shorter than the T_{PGM1} of the unselected cell. The difference is enough to program correctly. This means that this programming method can prevent write disturbance.

4. Conclusion

The OTP memories using standard CMOS devices that result high scalability are proposed. This programming method is effective for preventing write disturbance to unselected cells that share the gate line with a selected cell, because the gate poly-Si line break down after programming. In addition, a lower V_{PGM} is achieved without deep N-well (Table 1). Since this method does not require special manufacturing processes, it enables realization of small area and low-cost LSI with embedded one-time programmable memories.

References

[1] H. Ito, T. Namekawa, CICC 2004, pp.469-471.
[2] K. Matsufuji, et al., A-SSCC, pp. 212-215, 2007.
[3] M. Chen, et al., IEEE Electron Device Lett., vol.29, no. 5, pp.522-524, May. 2008.
[4] K. Zaitsu, et al., SSDM 2013, pp.698-699.

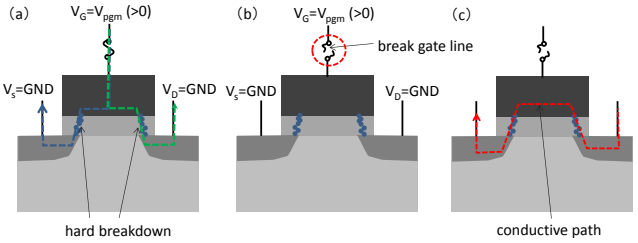

Fig. 1. One-time programming method in nMOSFET transistor.

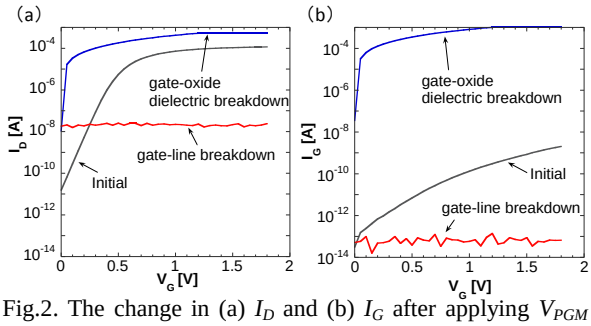

Fig.2. The change in (a) I_D and (b) I_G after applying V_{PGM} in NMOS transistor with gate length L=80 nm.

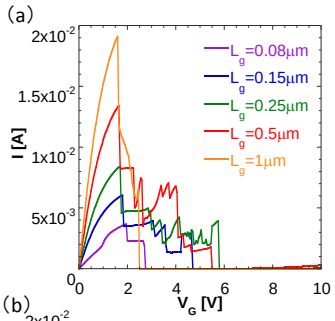

Fig.5. (a) Gate poly-Si line breakdown voltage dependence for various line widths. (b) Gate poly-Si line width L_g dependence of breakdown current I_{max}.

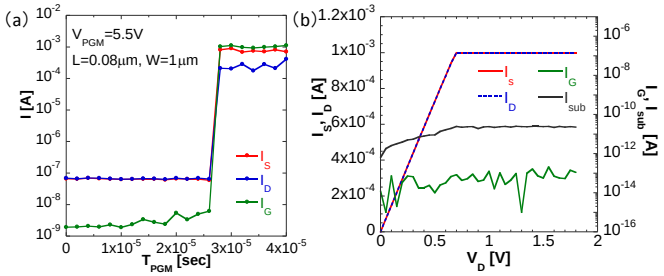

Fig.3. (a) Time dependence of I_S, I_D and I_G, (b) *I-V* characteristics of OTP memory after gate poly-Si line breakdown.

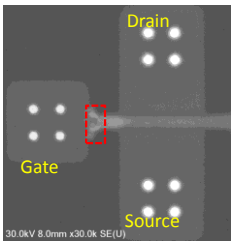

Fig.4. SEM of the anti-fuse OTP memory after applying V_{PGM}.

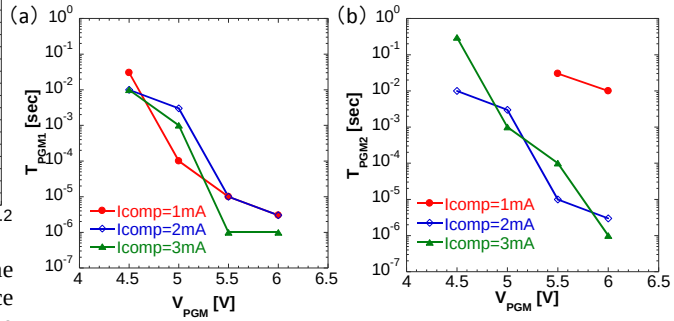

Fig.6. (a) Gate oxide breakdown voltage and (b) gate poly-Si line breakdown voltage dependences of programming time for various I_{comp}.

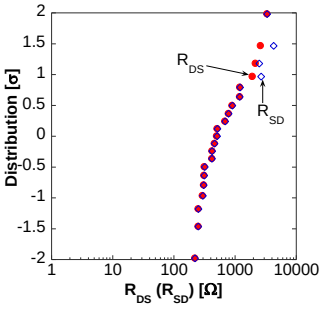

Fig.7. Distribution of R_{SD} (applying voltage at source) and R_{DS} (applying voltage at drain)

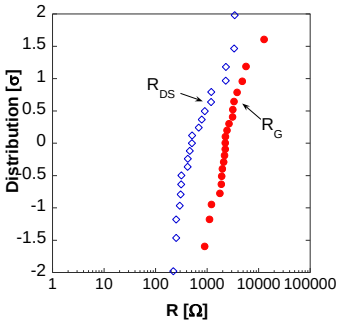

Fig.8. Resistance distribution after gate oxide breakdown and gate poly-Si line breakdown

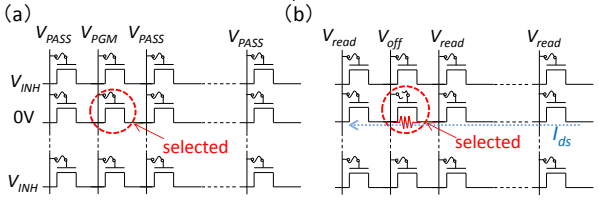

Fig.9. Example of cross-point OTP array.
(a) program operation, (b) read operation.

	L [nm]	V_{PGM} [V]	T_{PGM} [μ s]	Deep n-well
This work	50	5.5	10	Not Required
Ref.[4]	150	−13	400	Required

Table 1. The comparison between this work and the reference 4

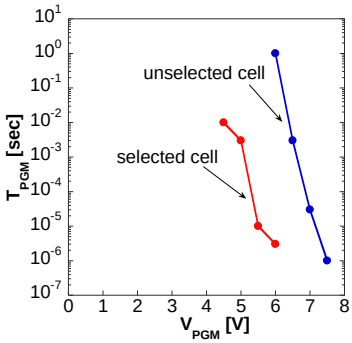

Fig.10. T_{PGM} of selected cell and unselected cell.