Shrinkable Triple Self-Aligned Field-Enhanced Split-Gate Flash Memory

Wen-Ting Chu, Hao-Hsiung Lin, Chia-Ta Hsieh, Hung-Cheng Sung, Yu-Hsiung Wang, Yung-Tao Lin, and Chung S. Wang

Abstract—This paper demonstrates a shrinkable triple self-aligned split-gate Flash cell fabricated using a standard 0.13- μm copper interconnect process. The approach used here to create a self-aligned structure is to form a spacer against the prior layer. Due to a higher aspect ratio when the cell pitch decreases, the profile of the spacer structure becomes sharper. This improves process control of the spacer profile and length. All the processes used here are compatible with standard logic process. The word line channel length of the cell is 0.11 μm . It is comparable in area with a stacked-gate cell and can be less than 13F^2 . The cell is erased by using poly-poly Fowler–Nordheim tunneling with a sharp floating-gate edge to increase the electric field, and is programmed by source-side injection. As a result, this cell is highly suitable for low power applications and embedded products. Characterization shows considerable program and erase speed, up to 300 K times cycling endurance, and excellent disturb margins.

Index Terms—Flash memory, self-aligned, source-side injection, split-gate.

I. INTRODUCTION

FLASH EEPROM, with its advantages of nonvolatility and in-system rewritability, is emerging as a mainstream memory technology. Apart from immunity from over-erase [1], [4], split-gate Flash EEPROM, as compared with its stacked-gate counterpart, can adopt a field-enhanced structure to further reduce its erase voltage [2], [3], [5], [6]. The cell is erased using Fowler–Nordheim (F–N) tunneling through interpoly oxide. In addition, source side hot electron injection is used to program the cell, which is at least one to two orders of magnitude larger in electron injection efficiency than the channel hot electron injection [8]–[10]. Therefore, its application to portable systems is highly promising due to the low power consumption during cell operation. In this paper, a triple self-aligned method [7] is demonstrated using a 0.13- μm standard logic process. As a result, the cell size can be shrunk to $0.214 \mu\text{m}^2$ ($\sim 13\text{F}^2$). The approach used to create a self-aligned structure was to form a spacer against the prior layer. The profile of the spacer structure becomes sharper as the cell pitch decreases, because the smaller pitch allows a higher aspect ratio with the same film thickness. This is an advantage

in forming a spacer with a sharper profile and allows better gate length control, which are the main difficulties faced in forming spacer structures. Therefore, the cell shrinkage becomes easier in the next generation and does not need expensive lithography and etching equipment. All the spacer forming processes are compatible with standard logic processes. The cell demonstrates an acceptable program and erase speed at voltages of 6 and 11 V, respectively, with quite a high immunity to program disturb, and several hundred thousand times program/erase (P/E) cycling endurance.

In the following sections, the device fabrication and cell operation are described, the cell program and erase speed, disturb performance and cycling endurance are presented. Finally, we will discuss the shrinkage potential for the next generation using this approach.

II. DEVICE FABRICATION

Drawings of the cross section of the cell and brief process flow are shown in Fig. 1. First, a floating gate (FG) oxide was grown, followed by a poly1 (as FG) and SiN deposition, then a standard shallow trench isolation (STI) process was implemented [Fig. 1(a)]. After STI formation, FG photo (defined as the “FG + source line + FG” area) and etch followed a thick SiN film deposition. A poly etch was used to produce a curved poly profile, creating a sharp FG structure [Fig. 1(b)] in the latter process. A tetraethoxylonesilane (TEOS) film was deposited and etched back to form a spacer, then a polyetch was performed with a TEOS spacer as a hard mask [Fig. 1(b)]. An oxide spacer for the source line (SL) poly and FG isolation was performed, then the source line poly, with the contact to the substrate, was deposited and etched back. Poly oxidation was performed to protect the SL poly. After the SiN strip, the exposed FG poly was removed. A high-temperature oxide (HTO), which serves as both the word-line (WL) gate oxide and the tunneling oxide (between the poly), was then deposited. Another poly was deposited to act as a WL gate. Finally, WL spacer formation and standard logic process followed [Fig. 1(c)]. The backend process employed Cobalt silicide and Cu interconnect to finish all cell and logic device fabrications. The schematic diagram of cell top-view and array configuration is shown in Fig. 1(d).

In Fig. 2(a), the cell cross section scanning electron microscope (SEM) picture shows that the FG and WL channel lengths are around 0.17 and 0.11 μm , respectively. The spacer profile is sharper than that of the 0.18- μm generation [7]. In addition, the transmission electron microscope (TEM) picture [Fig. 2(b)] shows that the sharp FG edge created by the poly etch is used to increase the electric field during poly–poly F–N tunneling when

Manuscript received May 12, 2004. The review of this paper was arranged by Editor R. Shrivastava.

W.-T. Chu and H.-H. Lin are with the Department of Electrical Engineering, Graduate Institute of Electronic Engineering, National Taiwan University, Taipei, Taiwan, R.O.C. (e-mail: wtchu@epicenter.ee.ntu.edu.tw).

C.-T. Hsieh, H.-C. Sung, Y.-H. Wang, Y.-T. Lin, and C. S. Wang are with the Nonvolatile Memory Division, Taiwan Semiconductor Manufacturing Company, Hsinchu, Taiwan 300, R.O.C.

Digital Object Identifier 10.1109/TED.2004.835995

Fig. 1. (a)–(c) Cross section delineations of the triple self-aligned cell process sequence. (d) The schematic diagram of cell top view and array configuration.

the cell is being erased. During program, the narrow gap between WL and FG will create a depletion region in the channel which offers a high electric field to accelerate the passing electrons, so the electrons will gain enough energy to overcome the oxide barrier and inject into FG.

III. CELL CHARACTERIZATION

Typical cell operation conditions are shown in Table I. Low voltage and low power are the main features of this cell. As can be seen in Fig. 3(a), cell current can be below $0.01 \mu\text{A}$ with a $10\text{-}\mu\text{s}$ programming time at source voltage of 6 V . We found that the higher the source voltage, the faster the program speed and the deeper the programming level (lower cell current). This is because a higher source voltage will create a larger electric

Fig. 2. (a) SEM picture of the cell in this paper. (b) TEM picture for a sharp FG corner (indicated by the circle) created by poly etch. An SiN film is capped here after the inter-poly oxide deposition to enhance the image contrast.

TABLE I
TYPICAL CELL OPERATION CONDITIONS

	Word line	Drain (Bit line)	Source
Program (Sel./Unsel.)	1.5V/0V	0.6V/2.5V	6V/0V
Erase	11V	0V	0V
Read-out	2.5V	1.5V	0V

field and at the same time couple the FG to a higher potential. The FG voltage (V_{FG}) can be expressed as follows:

$$V_{\text{FG}} = \frac{Q_{\text{FG}} + V_{\text{WL}} \times C_{\text{fw}} + V_{\text{sub}} \times C_{\text{fb}} + V_{\text{S}} \times C_{\text{fs}}}{C_{\text{fw}} + C_{\text{fb}} + C_{\text{fs}}}$$

where Q_{FG} is the charge on FG; C_{fw} , C_{fb} , and C_{fs} are the capacitances between FG and WL, substrate, and source, respectively, V_{FG} , V_{WL} , V_{sub} , and V_{S} are the voltages of FG, WL, substrate, and source, respectively. The source-coupling ratio here is around 70%. The reason why the source-coupling ratio still can be kept high as the cell shrinks is because the FG is self-aligned to the STI (there is no FG on the STI area), so the area of WL overlap to FG, i.e., C_{fw} , can be reduced. Compared with the nonself-aligned cell, in the self-aligned cell it is easier to shrink C_{fw} because of the overlay issue that is the main problem in split-gate cell technology. The erase speed is demonstrated in Fig. 3(b), with an erase voltage of 11 V applied to the WL. The cell current after a 10-ms voltage pulse erase is about $30 \mu\text{A}$, and can reach 90% of the current after a 100-ms voltage pulse erase. It means that the 10-ms erase time is very close to saturation.

Disturb is another crucial issue in Flash memory operations, and will become more serious as the cell pitch reduces. Punch-through, write, and read disturb are all well characterized here. Punch-through disturb occurs in the cell located on the cross

Fig. 3. (a) I_{cell} versus programming time for various source (V_S) and drain (V_D) voltage with 1.5 V WL voltage (V_{WL}). (b) I_{cell} versus erase time for various WL voltage.

Fig. 4. (a) Punch-through disturb: I_{cell} drop percentage versus program times (20 μs per pulse) with different V_D (bit line voltage at programming) at 25 $^{\circ}C$ and 125 $^{\circ}C$, for $V_{WL} = 0$ V, and $V_S = 6.5$ V. (b) Write disturb: I_{cell} drop percentage versus program times (20 μs per pulse) with different V_D (inhibited bit-line voltage at programming) at 25 $^{\circ}C$ and 125 $^{\circ}C$, for $V_{WL} = 1.5$ V, and $V_S = 6.5$ V.

of a selected bit-line (BL) and an unselected WL. The worst conditions for punch-through disturb are both high temperature and low drain voltage, as they both have a larger subthreshold

Fig. 5. Read disturb (soft erase): The time for I_{cell} change from programmed state to 1 μA versus reciprocal WL voltage ($1/V_{WL}$).

Fig. 6. Erased and programmed cell current as a function of P/E cycling times.

current, which will cause more hot electrons to be injected into the FG and induce a more serious cell current drop. As shown in Fig. 4(a), it can be seen that there is still enough margin to pass the program stress test at 125 $^{\circ}C$ seven times (for eight WL per page) with $V_D = 0.4$ V, and $V_{WL} = 0$ V. (Nominal V_D on the programmed cell is 0.6 V.)

Write disturb will occur in the cell located at the cross of a selected WL and an unselected BL. The worst conditions for write disturb are also high temperature and low drain voltage, as they both have a larger subthreshold current causing more hot electrons to be injected into the FG and inducing a more serious cell current drop. Fig. 4(b) shows that there is still a large enough margin to pass program stress test at 125 $^{\circ}C$ 4 K times (for a WL with 4 K cells) with 0.2 V back bias from drain to WL. The WL voltage on the programmed cell is 1.5 V, and the BL voltage on the unselected cell is 2.5 V (Table I).

As a field enhanced structure is employed to reduce erase voltage, read disturb (soft erase) is another issue which needs to be verified. Read disturb is characterized by erasing the cell with a different WL voltage. As shown in Fig. 5, by extrapolating the fitting line, when the WL voltage is 3.6 V, the time for the cell current to be changed from the programmed state to 1 μA is much longer than 10 years.

The cycling endurance performance illustrated in Fig. 6 shows that the cell current only drops by 28% after a 300 K times cycling stress test at room temperature, where the erase condition is $V_{WL} = 11$ V, 20 ms and the program condition is $V_S = 6$ V, 20 μs . In addition, there is no obvious current change for the programmed state.

Finally, the simulation result for the cell employed here using $V_{FG} = 7$ V and $V_S = 6$ V is shown in Fig. 7. During programming, the peak lateral and vertical electric fields are lo-

Fig. 7. Electric field simulation results where $V_S = 6$ V, $V_{FG} = 7$ V, $V_D = 0.6$ V, and $V_{WL} = 1.5$ V. Both vertical and lateral electric fields have a peak value around the gap region.

cated around the gap region, explaining why program efficiency is quite high even though the source voltage is only 6 V. In this simulation, $V_D = 0.6$ V and $V_{WL} = 1.5$ V.

IV. CONCLUSION

In this paper, we have demonstrated a shrinkable triple self-aligned cell employing poly-poly F-N tunneling and source side injection for Flash memory cell erase and program, respectively. A cell processed using the triple self-aligned method outlined here can be as small as $13F^2$ in area, which is comparable with stacked-gate technology. The process used for this cell is totally compatible with the standard logic process, so is suitable for embedded applications. The spacer structure here is very promising for cell shrink and would not be limited by lithography capabilities. In addition, the spacer structure resolves the mis-alignment issue that caused the split-gate cell to lose out to the stacked-gate cell. A poly etch is adopted in order to create a sharp corner to increase the tunneling oxide electric field, meaning that the poly-poly erase voltage can be reduced. The other approach to reducing the program power is to use source side injection to program the cell. A split-gate cell has the advantage of being free from over-erase compared to the stacked-gate cell, so the circuitry for this cell could be simple. Because of the shrinkable approach in cell manufacture, both low voltage and power consumption in cell operation, and simple circuitry, the cell described here is one of the best candidates for nonvolatile memory and embedded applications.

ACKNOWLEDGMENT

The authors would like to thank the members of the Non-volatile Memory Program, Research and Development Division, Taiwan Semiconductor Manufacturing Company, Ltd., for wafer fabrication and Silicon Storage Technology Inc., Sunnyvale, CA, for technical support.

REFERENCES

- [1] T. C. Ong, A. Fazio, N. Mielke, S. Pan, N. Righos, G. Atwood, and S. Lai, "Erratic erase in ETOX Flash memory array," in *Symp. VLSI Tech. Dig.*, 1993, pp. 83–84.
- [2] K.-C. Huang, Y.-K. Fang, D.-N. Yaung, C.-W. Chen, H.-C. Sung, D.-S. Kuo, and C. S. Wang, "The impacts of control gate voltage on the cycling endurance of split gate Flash memory," *IEEE Electron Device Lett.*, vol. 21, pp. 359–361, July 2000.
- [3] S. Kianian, A. Levi, D. Lee, and Y.-W. Hu, "A novel 3 volts-only, small sector erase, high density Flash E²PROM," in *Symp. VLSI Tech. Dig.*, 1994, pp. 71–72.
- [4] K. Naruke, S. Yamada, E. Obi, S. Taguchi, and M. Wada, "A new Flash-erase EEPROM cell with a select-gate on its source side," in *IEDM Tech. Dig.*, 1989, pp. 603–606.
- [5] B. Yeh, "Single transistor non-volatile electrically alterable semiconductor memory device," U.S. Patent 5 029 130, 1991.
- [6] C.-H. Wang, M. Hemming, P. Klinger, A. V. Kordes, C.-M. Liu, and K. Su, "On the cell misalignment for multilevel storage Flash E²PROM," in *Symp. VLSI Tech. Dig.*, 1999, pp. 191–194.
- [7] R. Mih, J. Harrington, K. Houlihan, H. K. Lee, K. Chan, J. Johnson, B. Chen, J. Yan, A. Schmidt, C. Gruensfelder, K. Kim, D. Shum, C. Lo, D. Lee, A. Levi, and C. Lam, "0.18- μ m modular triple self-aligned embedded split-gate Flash memory," in *Symp. VLSI Tech. Dig.*, 2000, pp. 120–121.
- [8] J. Van Houdt, P. Heremans, L. Deferm, G. Groeseneken, and H. E. Maes, "Analysis of the enhanced hot-electron injection in split-gate transistors useful EEPROM applications," *IEEE Trans. Electron Devices*, vol. 39, pp. 1150–1156, May 1992.
- [9] S. Bhattacharya, K. Lai, K. Fox, P. Chan, E. Worley, and U. Sharma, "Improved performance and reliability of split gate source-side injected Flash memory cells," in *IEDM Tech. Dig.*, 1996, pp. 339–342.
- [10] Y. Ma, C. S. Pang, J. Pathak, S. C. Tsao, C. F. Chang, Y. Yamauchi, and M. Yoshimi, "A novel high density contactless Flash memory array using split-gate source-side-injection cell for 5 V-only applications," in *Symp. VLSI Tech. Dig.*, 1994, pp. 49–50.

Wen-Ting Chu was born in Kaohsiung, Taiwan, R.O.C., in 1968. He received the B.S. degree in physics and the M.S. degree in electrical engineering both from National Taiwan University, Taipei, Taiwan, in 1991 and 1993, respectively. He is currently pursuing the Ph.D. degree at the Department of Electrical Engineering at the same university.

His research interests mainly focus on deep sub-micrometer nonvolatile memory technology development, Flash memory reliability studies, and MOS device physics. He joined the Taiwan Semiconductor Manufacturing Company in 1997, where he was primarily engaged in the development of nonvolatile memory, especially dedicated to embedded applications from the 0.5- μ m generation to the 0.13- μ m generation.

Hao-Hsiung Lin received the B.S., M.S., and Ph.D degrees in electrical engineering from National Taiwan University (NTU), Taipei, Taiwan, R.O.C. in 1978, 1980, and 1985, respectively.

He has been with the Department of Electrical Engineering, NTU since 1980, and is currently a Professor specializing in nanoelectronics. His current research interests include semiconductor materials and devices, especially on molecular beam epitaxy technology.

Chia-Ta Hsieh was born in Tainan, Taiwan, R.O.C. in 1965. He received his M.S. degrees in material science from the National Sun Yat-sen University, Kaohsiung, Taiwan, in 1992.

He has been with the Nonvolatile Memory Division, Taiwan Semiconductor Manufacturing Company (TSMC) since 1992. His research interest is mainly in the reliability of Flash memory. He currently leads the TSMC 0.13- μm Flash memory technology development project.

Yu-Hsiung Wang was born in Taichung, Taiwan, R.O.C. in 1972. He received the M.S. degree in electrical engineering in 1996 from National Sun-Yat Sen University, Kaohsiung, Taiwan. He is currently pursuing the Ph.D. degree in electrical engineering at National Tsing Hua University, Hsinchu, Taiwan for the modeling and reliability of Flash memory.

He joined the Nonvolatile Memory Division, Taiwan Semiconductor Manufacturing Company in 1998. His research interests and activities cover hot-carrier effects, gate oxide integrity, electrical characterization techniques, and modeling and simulation of Flash memory.

Yung-Tao Lin, photograph and biography not available at the time of publication.

Hung-Cheng Sung, photograph and biography not available at the time of publication.

Chung S. Wang, photograph and biography not available at the time of publication.