



DECLARATION OF GORDON MACPHERSON

I, Gordon MacPherson, am over twenty-one (21) years of age. I have never been convicted of a felony, and I am fully competent to make this declaration. I declare the following to be true to the best of my knowledge, information, and belief:

1. I am the Director, Board Governance & Policy Development at The Institute of Electrical and Electronics Engineers, Incorporated ("IEEE").
2. IEEE is a neutral third party in this dispute.
3. I am not being compensated for this declaration and IEEE is only being reimbursed for the cost of the article I am certifying.
4. Among my responsibilities as Director of Board Governance & Policy Development, I act as a custodian of certain records for IEEE.
5. I make this declaration based on my personal knowledge and information contained in the business records of IEEE.
6. As part of its ordinary course of business, IEEE publishes and makes available technical articles and standards. These publications are made available for public download through the IEEE digital library, IEEE Xplore.
7. It is the regular practice of IEEE to publish articles and other writings including article abstracts and make them available to the public through IEEE Xplore. IEEE maintains copies of publications in the ordinary course of its regularly conducted activities.
8. The articles below have been attached as Exhibits A-D to this declaration:

Exhibit A	F. Zhou et al., "VDNROM: A Novel Four-Bits-Per-Cell Vertical Channel Dual-Nitride-Trapping-Layer ROM for High Density Flash Memory Applications," 2006 European Solid-State Device Research Conference, Montreux, Switzerland, 2006, pp. 226-229, doi: 10.1109/ESSDER.2006.307679.
Exhibit B	E.-K. Lai et al., "A Multi-Layer Stackable Thin-Film Transistor (TFT) NAND-Type Flash Memory," 2006 International Electron Devices Meeting, San Francisco, CA, USA, 2006, pp. 1-4, doi: 10.1109/IEDM.2006.346903.
Exhibit C	A. Arreghini et al., "Experimental Characterization of the Vertical Position of the Trapped Charge in Si Nitride-Based Nonvolatile Memory Cells," in IEEE Transactions on Electron Devices, Vol. 55, No. 5, pp. 1211-1219, May 2008, doi: 10.1109/TED.2008.919713.

Exhibit D	H. Tanaka <i>et al.</i> , "Bit Cost Scalable Technology with Punch and Plug Process for Ultra High Density Flash Memory," <i>2007 IEEE Symposium on VLSI Technology</i> , Kyoto, Japan, 2007, pp. 14-15, doi: 10.1109/VLSIT.2007.4339708.
-----------	---

9. I obtained a copy of Exhibit A through IEEE Xplore, where it is maintained in the ordinary course of IEEE's business. Exhibit A is a true and correct copy of the Exhibit, as it existed on or about August 5, 2025.
10. I obtained a copy of Exhibit B through IEEE Xplore, where it is maintained in the ordinary course of IEEE's business. Exhibit B is a true and correct copy of the Exhibit, as it existed on or about August 5, 2025.
11. I obtained a copy of Exhibit C through IEEE Xplore, where it is maintained in the ordinary course of IEEE's business. Exhibit C is a true and correct copy of the Exhibit, as it existed on or about August 5, 2025.
12. I obtained a copy of Exhibit D through IEEE Xplore, where it is maintained in the ordinary course of IEEE's business. Exhibit D is a true and correct copy of the Exhibit, as it existed on or about August 6, 2025.
13. The articles and abstracts from IEEE Xplore show the date of publication. IEEE Xplore populates this information using the metadata associated with the publication.
14. F. Zhou *et al.*, "VDNROM: A Novel Four-Bits-Per-Cell Vertical Channel Dual-Nitride-Trapping-Layer ROM for High Density Flash Memory Applications," was published in 2006 European Solid-State Device Research Conference, held on September 19-21, 2006, in Montreux, Switzerland. Copies of the conference proceedings were made available no later than the last day of conference. This article was added to the IEEE digital library Xplore on February 12, 2007, as shown from the abstract from IEEE Xplore, and was publicly available for download from IEEE Xplore by February 12, 2007, and is currently available for public download from the IEEE digital library, IEEE Xplore.
15. E -K. Lai *et al.*, "A Multi-Layer Stackable Thin-Film Transistor (TFT) NAND-Type Flash Memory," was published in 2006 International Electron Devices Meeting, held on December 11-13, 2006, in San Francisco, CA, USA. Copies of the conference proceedings were made available no later than the last day of conference. This article was added to the IEEE digital library Xplore on April 16, 2007, as shown from the abstract from IEEE Xplore, and was publicly available for download from IEEE Xplore by April 16, 2007, and is currently available for public download from the IEEE digital library, IEEE Xplore.
16. A. Arregghini *et al.*, "Experimental Characterization of the Vertical Position of the Trapped Charge in Si Nitride-Based Nonvolatile Memory Cells," was published in IEEE Transactions on Electron Devices Journal, (Volume: 55, Issue: 5, pages 1211-1219, dated as of May 31, 2008). This article is currently available for public download from the IEEE digital library, IEEE Xplore.
17. H. Tanaka *et al.*, "Bit Cost Scalable Technology with Punch and Plug Process for Ultra High Density Flash Memory," was published in 2007 IEEE Symposium on VLSI

Technology, held on June 12-14, 2007, in Kyoto, Japan. Copies of the conference proceedings were made available no later than the last day of conference. This article was added to the IEEE digital library Xplore on October 8, 2007, as shown from the abstract from IEEE Xplore, and was publicly available for download from IEEE Xplore by October 8, 2007, and is currently available for public download from the IEEE digital library, IEEE Xplore.

I hereby declare that all statements made herein of my own knowledge are true and that all statements made on information and belief are believed to be true, and further that these statements were made with the knowledge that willful false statements and the like are punishable by fine or imprisonment, or both, under 18 U.S.C. § 1001.

I declare under penalty of perjury that the foregoing statements are true and correct.

Date: 8/7/2025

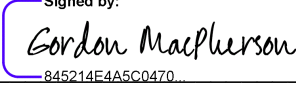
By:  Signed by:
Gordon Macpherson
845214E4A5C0470

EXHIBIT A

VDNROM: A Novel Four-Bits-Per-Cell Vertical Channel Dual-Nitride-Trapping-Layer ROM for High Density Flash Memory Applications

Falong Zhou, Yimao Cai, Ru Huang, Yan Li, Xiaonan Shan, Jia Liu,
Ao Guo, Xing Zhang, and Yangyuan Wang

Institute of Microelectronics, Peking University, Beijing 100871, China
Tel: 86-10-62757761, Fax: 86-10-62751789
Email: huangr@ime.pku.edu.cn

Abstract - A novel vertical channel nonvolatile memory cell with oxide-nitride-oxide-nitride-oxide (ONONO, dual nitride trapping layers) dielectrics stack is proposed and experimentally demonstrated for the first time. Compared with the conventional planar NROM cell, since the cell area of the proposed vertical structure is independent of the gate length, the VDNROM structure can relax the limitation of the gate length scaling, and can have high capability of cell area shrinking. The fabrication process of this VDNROM device is basically compatible with planar CMOS technology. The VDNROM cell can be programmed and erased by the hot carrier injection to the localized trapping dual-nitride-layers, so it can achieve a four physical bits storage capability each cell. The reliability behaviors including the cycling endurance and the bake retention at 150°C have also been investigated and show the acceptable characteristics. The experiment results verify the VDNROM cell as a good candidate for high-density applications.

I. INTRODUCTION

Recently, SONOS flash memory has been paid more attention and considered as a promising technology bridging the gap between the floating-gate flash and the next generation memories because of its intrinsic better endurance, good scalability, simpler fabrication process and its perfect compatibility with conventional CMOS technology for embedded applications [1-3]. However, conventional SONOS cell is only one-bit-per-cell storage, and the data retention is still a big problem due to the direct tunneling of stored charge through the ultra thin tunneling oxide.

The NROM concept [4-5] was proposed to solve the drawbacks of SONOS by using channel hot electrons injection (CHEI) for programming and band-to-band tunneling induced hot holes injection (BBTIHH) for erasing. This device can allow thicker tunneling oxide, while still maintaining high efficient Program /Erase (P/E) operations. Moreover, the P/E

mechanisms lead to localized charge storage in the trapping nitride layer, so the NROM can store two distinct bits per cell. However, the further scaling of the cell area of the conventional planar NROM device for higher density is limited by the gate length scaling for two reasons. Firstly, the drain voltage can not be scaled below 3.2V for CHEI, and therefore the punch-through effect would become worse as the gate length scaling [6]. Secondly, since two bits are stored in physically distinct areas in the nitride layer, the lateral distributions of the charges have to be well separated to avoid any interference between the two bits [7].

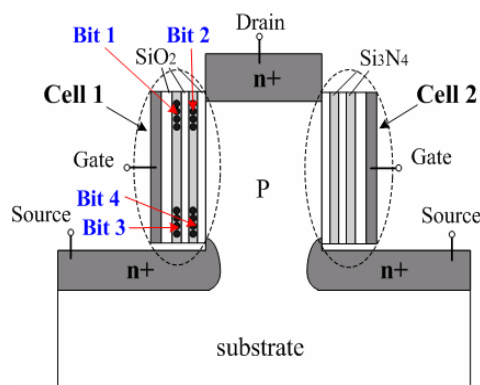


Fig. 1 Schematic cross sectional view of the proposed 4 bits/cell VDNROM flash memory.

In this work, as shown in Fig. 1, we proposed and demonstrated experimentally a novel vertical channel dual-nitride-trapping-layers ROM (VDNROM) structure nonvolatile memory device achieving four-bits-per-cell function for the first time. Compared with the conventional planar NROM cell, since the cell area of this vertical channel structure is independent of the gate length, the VDNROM structure can relax the limitation of the gate length scaling, and can have high capability to shrink the cell area. Moreover, the fabrication of this VDNROM device is basically compatible with planar

CMOS technology. The experimental results demonstrate continued better scalability of the VDNROM cell for high-density applications.

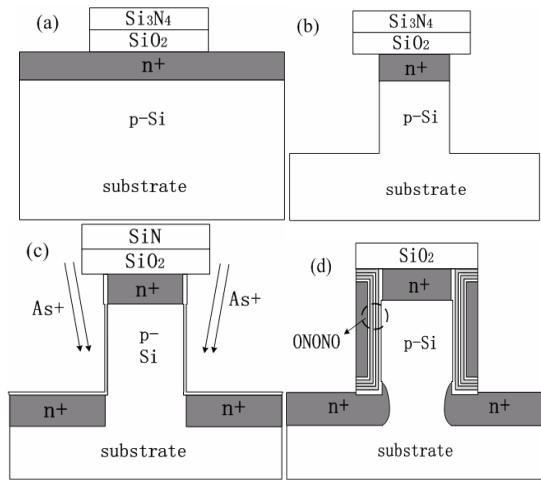


Fig. 2. The process flow of the proposed 4bits/cell VDNROM flash memory.

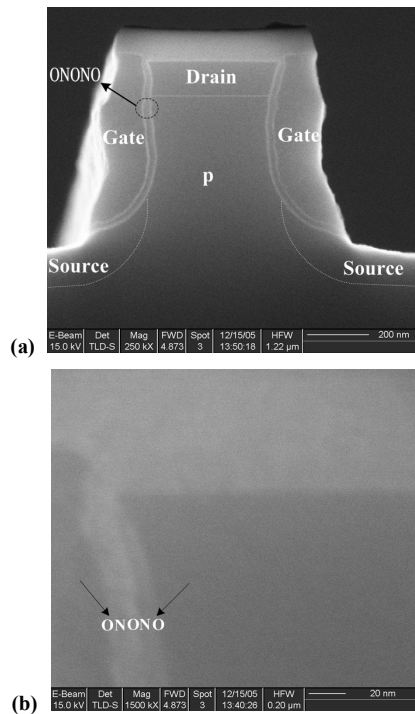


Fig. 3. The SEM images of the proposed 4bits/cell VDNROM flash memory cell (a) and the ONONO gate dielectric stack (b).

II. DEVICE STRUCTURE AND FABRICATION

The schematic cross-sectional view of the proposed VDNROM flash memory cell with the oxide-nitride-oxide-nitride-oxide (ONONO) gate dielectric stack is illustrated in Fig. 1. The charges can be stored in the

dual nitride trapping layers above the channel next to the n+/p junctions, so this VDNROM cell can achieve a four physical bits storage capability as shown in Fig. 1. The tunneling oxide and the block oxide are thicker than 5 nm to avoid any direct tunneling of the charges, and the gate dielectric stack of the ONONO five layers (the tunneling oxide, the first trapping nitride layer, the inter-nitride oxide, the second trapping nitride layer, the block oxide) are 7/4/5/5/13 nm respectively.

Referring to Fig. 1, the process flow to realize such a 4bits/cell VDNROM device was shown in Fig. 2(a)-(b). After isolation, the channel boron doping and the drain arsenic doping were implemented. Two layers of oxide and nitride was deposited, patterned and etched to form the dual-layers hard masks as shown in Fig. 2(a). The silicon substrate was then etched by ICP (Inductive Coupled Plasma) to form the steep device mesa. Next, the device mesa was etched isotropically by RIE (Reactive Ion Etch) to form the recessed self-aligned poly gate structure shown in Fig. 2(b). To remove the surface damage on the sidewalls of the device mesa, thin sacrificial oxide was grown, which was followed by source arsenic implantation ($4 \times 10^{15} \text{ cm}^{-2}$, 100KeV), as shown in Fig. 2(c). After sacrificial oxide removed, 7.0-nm gate oxide was grown at 850°C by dry thermal oxidation. Next, a dielectric stack of trapping-nitride/inter-oxide/trapping-nitride/block-oxide (NONO) with 4/5/5/13 nm respectively was deposited. Then, a polysilicon film was deposited, implanted and etched to form the self-aligned gate spacer. After the nitride hard mask removed, rapid thermal annealing was implemented to activate the dopants in the S/D region and poly gate, as shown in Fig. 2(d). Finally, the device was covered by low temperature deposited oxide (LTO), and followed by the rest of the process such as contact opening, metal defining and annealing. Fig. 3 (a) shows the SEM (scanning electron microscope) micrograph of the cross section of the fabricated VDNROM cell, and the SEM image of the ONONO gate dielectric stack is shown in Fig. 3 (b).

III. OPERATION MECHANISM AND EXPERIMENT RESULTS

The VDNROM cell is programmed and erased by channel hot carriers injection. The four states (“00”, “01”, “10”, “11”) corresponding to bit-1 and bit-2 are shown in Fig. 4 (a)-(d), and the programming and erasing operations of the four states are shown in Table I. For example, applying 7.0 V on the drain electrode, 12.0 V on WL (the gate electrode) and grounding the source electrode for programming the state “00” in Fig. 4(a). The accelerated channel electrons are injected into the both two nitride layers of the ONONO gate dielectric stack near the drain junction, a self-aligned storage with the junction edge. As shown in Fig. 4(b) and Table I, the state of “01” is programmed by the hot electrons injection to the first trapping nitride layer with lower drain voltage 5.0 V (compared with the programming of the state of “00”), 6.0 V on WL (the

gate electrode) and grounding the source electrode. The state of “10” shown in Fig. 4(c) is realized by two steps. As shown in Table I, the first step is the hot electrons injection to the dual trapping nitride layers with $V_g=12.0$ V, $V_d=7.0$ V and $V_s=0$ V, and the second is the hot holes injection to the first nitride layer applying $V_d=-8.0$ V, $V_g=5.0$ V and $V_s=0$ V. Holes are generated by band-to-band tunneling in the deep depletion layer under the gate dielectric stack above the n+/p junction [4, 8]. Holes increase their energy due to the lateral field in depletion region of the channel and they are injected through the bottom oxide into the nitride trapping layer, where they recombine with the stored electrons or are trapped. As shown in Fig. 4(d), the state of “11” is the initial state and it can also be realized by erasing from other states shown in Table I.

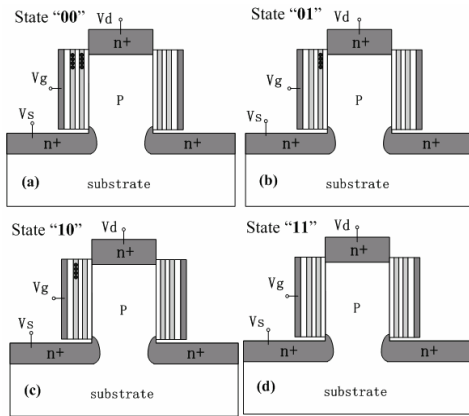


Fig. 4. The four states corresponding to bit-1 and bit-2 of the proposed VDNROM flash memory.

TABLE I.
VDNROM Flash Memory Program/Erase Operating Conditions
(Four states corresponding to bit-1 and bit-2)

State	P/E	Gate (Pulse)	Drain (Pulse)	Source	Time	V _{th}
'00'	Program from "11"	12V 1μs 0V 9μs	7V 1μs 0V 9μs	0 V	10 ms	5.5V
'01'	Program from "11"	6V 1μs 0V 9μs	5V 1μs 0V 9μs	0 V	10 ms	4.3V
'10'	Program from "11"	12V 1μs 0V 9μs -8V 9μs	7V 1μs 0V 9μs 5V 9μs	0 V	100 ms	3.3V
'11'	Erase	0V 1μs -12V 1μs	6V 1μs 0V 9μs	0 V	100 ms	2.3V

Fig. 5 shows transfer characteristics of the VDNROM flash memory cell with 0.3μm channel length in the four states. The threshold voltage of State “11” is 2.3V (the same V_{th} as the initial). State “00” has a threshold shift of $\Delta V_{th}=3.2$ V from State “11”. The ΔV_{th} of 2.0V in State “01” is observed. The threshold voltage of State “10” is 3.3V, and the

ΔV_{th} is 1.0V. The transfer characteristics of each state exhibits excellent subthreshold performance as shown in Fig. 5. The programming and erasing characteristics of the VDNROM flash memory are shown in Fig. 6.

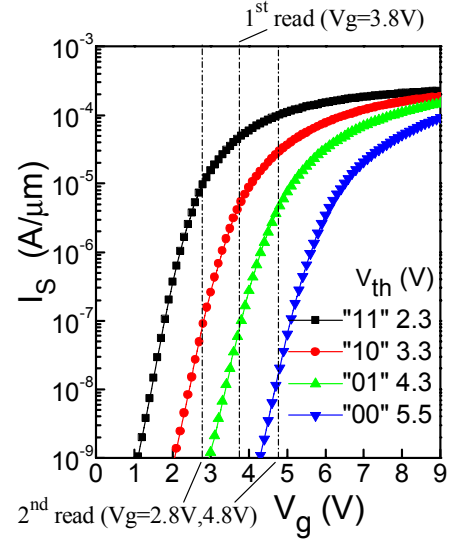


Fig. 5. The transfer characteristics of the programmed/erased bit-1 and bit-2 of the VDNROM flash memory.

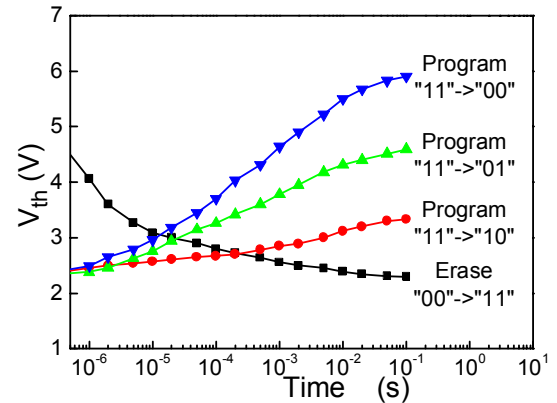


Fig. 6. The Programming/Erasing characteristics of the VDNROM flash memory.

The device is read in the “reverse” direction compared to programming, by applying the read biases to source electrode (1.5V) and grounding the drain electrode. This reverse read mode can enhance the threshold voltage sensitivity to the localized trapping charges [4]. As shown in Fig. 5, during the first read operation, the “00” and “01” states can be distinguished from the “10” and “11” states at 3.8V (gate voltage). After that, State “11” from State “10” and State “01” from State “00” at 2.8V and 4.8V (gate voltage), respectively, is performed for the second read operation.

Similarly, the bit-3 and bit-4 can be programmed, erased and read by interchanging the role of the drain/source electrodes.

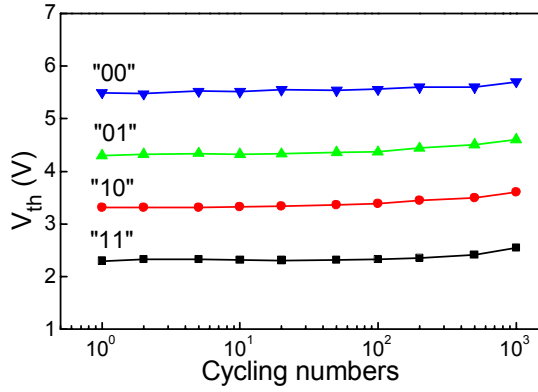


Fig. 7. The Endurance characteristics of the VDNROM flash memory.

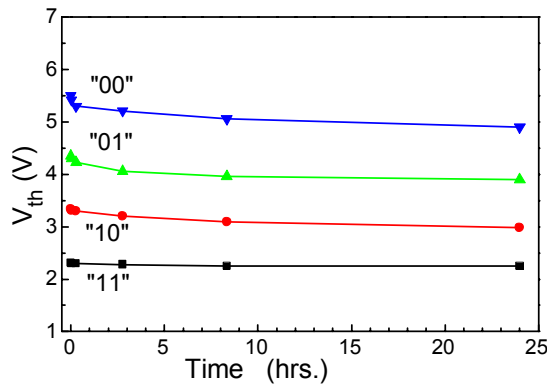


Fig. 8. The bake retention performance of the VDNROM flash memory at 150°C

Fig. 7 shows the endurance characteristics of the four states (corresponding to bit-1 and bit-2). The 2-bit cycling loop consists of the sequence of bit-2 program (State "11" to State "01") → bit-1 program ("01" to "00") → bit-2 erase ("00" to "10") → bit-1 erase ("10" to "11") that can be considered as the worst case, because the tunneling oxide has been tunneled by twice hot electrons and twice hot holes in one cycle (the inter-nitride oxide by once electrons and once holes). After 10³ P/E cycles, the worst Vth window of 0.8V is maintained in spite of the Vth increase at the erased cell. These electrical results indicate that the cycling endurance performance of the VDNROM cell is acceptable.

Fig. 8 shows the bake retention performance of the VDNROM measured at 150°C. 0.4V small degradation and saturation of Vth are observed after 24 hours baking measurement. With the better quality tunneling oxide, the smaller degradation of Vth window can be observed.

IV. CONCLUSION

For the first time, a novel VDNROM flash memory cell with an oxide-nitride-oxide-nitride-oxide (ONONO) gate dielectric stack has been proposed, demonstrated experimentally, and successfully operated in four-bits-per-cell mode in the 0.3μm channel length vertical device. In the conventional planar NROM cell, the cell area further scaling is limited by the gate length scaling. In contrast, since the cell area of the VDNROM structure is independent of the gate length, the structure can have high capability of the cell area shrinking with the relaxed requirement of the gate length scaling. Moreover, the fabrication of the VDNROM is basically compatible with planar CMOS technology. The reliability performances including the cycling endurance and the bake retention at 150°C have also been investigated and are acceptable. This new technology can be used for data storage due to the high density, and for embedded applications due to the simplicity of the process.

ACKNOWLEDGMENT

The authors would like to thank the Microfabrication Laboratory staffs of Peking University for their supports in device fabrication. This work is supported by National Natural Science Foundation of China (No. 90207004) and State Key Fundamental Research Project of China.

REFERENCES

- [1] Y. Kamigaki, S. Minami, "MNOS nonvolatile semiconductor memory technology: Present and future," *IEICE Trans. Electron.*, vol. E84-C, no. 6, Jun. 2001.
- [2] M. H. White, D.A. Adams, J. Bu, "On the go with SONOS," *IEEE Circuits Devices Mag.*, no. 7, pp. 22–31, Jul. 2000.
- [3] Peiqi Xuan, Min She, B. Harteneck, A. Liddle, J. Bokor, T.J. King, "FinFET SONOS flash memory for embedded applications," in *Technical Digest of IEDM*, pp. 26.4.1–26.4.4, Dec. 2003.
- [4] B. Eitan, P. Pavan, I. Bloom, A. Efraim, A. Frommer, and D. Finzi, "NROM: a novel localized trapping 2-bit nonvolatile memory cell," *IEEE Electron Device Lett.*, vol. 21, no. 6, pp. 543–545, Jun. 2000.
- [5] E. Mayaan, R. Dvir, J. Shor, Y. Polansky, Y. Sofer, I. Bloom, D. Avni, B. Eitan, Z. Cohen, M. Meyassed, Y. Alpern, H. Palm, E. Stein, V. Kamienski, P. Haibach, D. Caspary, S. Riedel, and R. Knöfler, "A 512Mb NROM flash data storage memory with 8MB/s data rate," in *Int. Solid State and Circuits Conf.*, pp. 100–101, 2002.
- [6] L. Breuil, L. Haspeslagh, P. Blomme, D. Wellekens, J. De Vos, M. Lorenzini, J. Van Houdt, "A new scalable self-aligned dual-bit split-gate charge-trapping memory device," *IEEE Trans. Electron Devices*, vol. 52, Issue 10, pp. 2250 – 2257, Oct. 2005.
- [7] H. T. Lu, T. H. Hsu, M. T. Wu, K. Y. Hsieh, R. Liu, C. Y. Lu, "Studies of the reverse read method and second-bit effect of 2-bit/cell nitride-trapping device by quasi-two-dimensional model," *IEEE Trans. Electron Devices*, vol. 53, Issue.1, pp.119–125, Jan. 2006.
- [8] S. Wolf, "Silicon processing for the VLSI era," in *The Submicron MOSFET*. Sunset Beach, CA: Lattice, p. 199, 1994.

EXHIBIT B

A Multi-Layer Stackable Thin-Film Transistor (TFT) NAND-Type Flash Memory

*Erh-Kun Lai, Hang-Ting Lue, Yi-Hsuan Hsiao, Jung-Yu Hsieh, Chi-Pin Lu, *Szu-Yu Wang, Ling-Wu Yang, Tahone Yang, Kuang-Chao Chen, †Jeng Gong, Kuang-Yeu Hsieh, Rich Liu, and Chih-Yuan Lu

Macronix International Co. Ltd., 16, Li-Hsin Road, Hsinchu Science Park, Hsinchu, Taiwan, R.O.C.

†Institute of Electronics Engineering, National Tsing Hua University, Hsinchu, Taiwan, R.O.C.

*Also with Institute of Electronics Engineering, National Tsing Hua University, Hsinchu, Taiwan, R.O.C.

E-mail: eklai@mxic.com.tw, Fax: 886-3-5789087; Phone: 886-3-5786688

Abstract

A double-layer TFT NAND-type Flash memory is demonstrated, ushering into the era of three-dimensional (3D) Flash memory. A TFT device using bandgap engineered SONOS (BE-SONOS) [1,2] with fully-depleted (FD) poly silicon (60 nm) channel and tri-gate P⁺-poly gate is integrated into a NAND array. Small devices (L/W=0.2/0.09 μm) with excellent performance and reliability properties are achieved. The bottom layer shows no sign of reliability degradation compared to the top layer, indicating the potential for further multi-layer stacking. The present work illustrates the feasibility of 3D Flash memory.

I. Introduction

Recently, the NAND flash becomes very popular for portable electronics, and the demand for memory density multiplies every year. Although NAND Flash is aggressively scaled [3,4], it becomes more and more difficult to follow Moore's law to continue the scaling due to both process and device limitations. Three-dimensional (3D) multi-layer-stack memory is another path to ultra-high density memory.

We demonstrated a low-thermal budget TFT NAND device recently [2]. The TFT memory employs BE-SONOS [1] device, which provides both efficient hole tunneling erase and good data retention [1]. Moreover, BE-SONOS uses only matured front-end materials (oxide, nitride and poly gate). This greatly enables the multi-layer stack capability without metal contamination issues. In this work, we further fabricate a double-layer TFT NAND Flash. The performance and reliability of bottom and top layers are examined to provide guidelines for further multi-layer stacking. The effects of device dimension, reliability, and NAND array operation are addressed.

II. Fabrication

The typical device parameters and fabrication flow are shown in Table 1 and Table 2. Figures 1-4 show the cross-sectional views of the double layer TFT NAND devices. The TFT device has a 60 nm thick poly silicon channel over the buried oxide. The poly silicon is an amorphous silicon (a-Si) layer deposited by low-pressure chemical vapor deposition

(LPCVD) process, followed by a low-temperature thermal annealing (600 °C) for crystallization. A multi-layer O1/N1/O2 tunneling dielectric is used to achieve easy hole tunneling during erase, while eliminating the direct tunneling leakage during retention [1]. Next, SiN trapping layer (N2) and top blocking oxide (O3) are grown. A heavily doped P⁺-poly gate is adopted to suppress the gate injection during -FN erase. The multi-layer ONONO are shown in Fig. 1. The device has a tri-gate structure as shown in Fig. 2.

The top oxide process has the largest thermal budget. Two top oxide (O3) formation processes are compared: an LPCVD oxide (HTO) with rapid thermal annealing, and an in-situ steam generation (ISSG) oxidation to convert a portion of the trapping nitride (N2) into oxide.

After gate patterning, shallow junction is formed by low-energy As⁺ implantation. Planarization is then carried out by HDP oxide deposition and CMP. After forming the bottom-layer TFT devices, the processes are duplicated to form the top-layer TFT devices. Contact etching for the two layers are performed independently to avoid over etching. The final double-layer TFT NAND Flash memory is shown in Figs. 3-4.

III. Device Characterization

(a) Short-Channel Effect:

The I - V and V_T for the memory cells with various channel lengths are shown in Figs. 5, 6, respectively. Similar to bulk devices, the transconductance (g_m) increases with shorter channel lengths. Figure 6 shows that the -FN erased state (-17 V, 50 msec) follows the same V_t -roll-off characteristics with the initial state, but with 3 V lower V_t . ISSG and HTO top oxide devices show similar V_t roll-off trends.

(b) Narrow-Width Effect:

The I - V characteristics for memory cells with various channel widths are shown in Fig. 7. Generally a narrower device shows a better subthreshold slope (S.S.). Figure 8 shows that for a narrow device the transconductance per width (g_m/W) increases, while V_t decreases. Thus a narrower device has a better DC performance due to better gate control capability of the tri-gate structure.

(c) Program/Erase Characteristics:

The device is programmed by +FN electron injection (Fig. 9), and erased by -FN hole tunneling (Fig. 10). P⁺-poly gate is used to suppress the gate electron injection during -FN erase. Accordingly, the device can be erased into depletion mode. Both +FN program and -FN erase saturate for each bias voltage. Samples with ISSG and HTO top oxide show similar P/E speed.

(d) P/E Cycling Endurance:

The P/E cycling endurance is shown in Fig. 11. Erase degradation occurs at < 1K dumb-mode cycles. ISSG top oxide devices show better endurance than those with HTO top oxide. The top oxide process has large impact on the cell endurance because top oxide suffers a large electric field stress during P/E, while the top oxide has lower quality due to the lower thermal process.

The endurance of TFT device is understandably inferior to the bulk BE-SONOS device [1,5], where a high-temperature (950~1000 °C) process is used to convert nitride into top oxide. The reason may be attributed to that the lower temperature ISSG and HTO top oxides have lower quality that easily trap charges after P/E cycle stressing, resulting in lower endurance.

Adjusted P/E cycling (as commonly practiced) achieves better endurance (Fig. 12). The corresponding P/E shot numbers are recorded in Fig. 13. After 1K cycling the device starts to degrade and more P/E shots are needed to maintain the V_t levels. Figure 14 illustrates that ISSG top oxide has a better subthreshold slope than HTO top oxide. Typical *I/V* characteristics during 10K P/E cycling are shown in Fig. 15.

IV. Comparison between the Top and Bottom Layers

Figure 16 compares the V_T-roll-off of erase saturation state for the bottom and top layers. The bottom layer has a slightly faster V_T-roll-off. In addition, the V_T distribution of the bottom-layer is wider. This is because the bottom layer undergoes a higher thermal budget. For better bottom layer performance thermal budget should be further reduced.

The P/E cycle endurance, S.S. degradation, and retention are shown in Figs. 17, 18 and 19, respectively, for the short-channel device (L_g/W=0.2/0.09 μm). The bottom layer shows no sign of reliability degradation compared to the top layer. This indicates that further multi-layer stack is feasible.

V. NAND Array Functions

The reading method for the NAND array is shown in Fig. 20. Since TFT NAND array has higher series resistance compared to the bulk device, read current saturates at large gate overdrive. The read current is enhanced by increasing V_{PASS}. For V_{PASS}=7 V, and V_{READ}=1 V, the read current for a 16-string NAND array approaches 1 μA, ample for data Flash design.

The program-inhibit characteristics of TFT NAND array are illustrated in Fig. 21. Cells C and D are under medium-field gate disturb, while cell B can be program-inhibited by raising the channel potential.

VI. Summary

For the first time, a successful double-layer TFT NAND Flash memory is demonstrated. The experimental results are very encouraging for future 3D Flash memories.

VII. Acknowledgements

Special thanks to Dr. R.I. Huang, Dr. S.P. Hong, Martin Wu, Dr. Chris Luoh, Terence Hung, C.F. Chen, Bear Wu and Tammy Lin, Technology Development Center in Macronix Inc., for the important cooperation and contributions to this work.

References

References:

- [1] H.T. Lue, S.Y. Wang, E.K. Lai, Y.H. Shih, S.C. Lai, L.W. Yang, K.C. Chen, J. Ku, K.Y. Hsieh, R. Liu, and C.Y. Lu, "BE-SONOS: A Bandgap Engineered SONOS with Excellent Performance and Reliability", *Tech. Digest 2005 International Electron Devices Meeting*, pp. 555-558, 2005.
- [2] E.K. Lai, H.T. Lue, Y.H. Hsiao, J.Y. Hsieh, S.C. Lee, C.P. Lu, S.Y. Wang, L.W. Yang, K.C. Chen, J. Gong, K.Y. Hsieh, J.Ku, R. Liu, and C.Y. Lu, "A Highly Stackable Thin-Film Transistor (TFT) NAND-Type Flash Memory", *VLSI Technology Symposium*, pp. 56-57, 2006.
- [3] K. T. Park, J. Choi, J. Sel, V. Kim, C. Kang, Y. Shin, U. Roh, J. Park, J. S. Lee, J. Sim, S. Jeon, C. Lee and K. Kim, "A 64-Cell NAND Flash Memory with Asymmetric S/D Structure for Sub-40nm Technology and Beyond", *VLSI Technology Symposium*, pp. 24-25, 2006.
- [4] K. Kim, "Technology for sub 50 nm node DRAM and NAND Flash Manufacturing", *Tech. Digest 2005 International Electron Devices Meeting*, pp. 333-336, 2005.
- [5] H. T. Lue, S. Y. Wang, Y. H. Hsiao, E. K. Lai, L. W. Yang, T. Yang, K. C. Chen, K. Y. Hsieh, R. Liu, and C. Y. Lu, "Reliability Model of Bandgap Engineered SONOS (BE-SONOS)" *Tech. Digest 2006 International Electron Devices Meeting*, session 18-4, 2006. (this conference).

Poly channel thickness	60 nm
O1	1.5 nm (ISSG)
N1	2 nm (LPCVD)
O2	2 nm (LPCVD)
N2	7 nm (LPCVD)
O3	9 nm (ISSG or HTO)
Poly gate	P ⁺ -poly
Channel length	0.2 μm
Channel width	0.09 μm

Table 1 Typical process condition of TFT NAND device.

1. Si substrate	9. N ⁺ S/D & pocket implantation
2. Buried oxide	10. HDP CVD deposition
3. a-Si channel deposition	11. Oxide CMP planarization
4. Low-temperature annealing	12. Repeat processes from a-Si to N ⁺ S/D implantation
5. Channel patterning	13. ILD deposition and planarization
6. P-type well doping	14. Contact patterning
7. BE-SONOS: ONONO formation	15. Metal deposition and patterning
8. P ⁺ poly gate patterning	16. Alloy annealing

Table 2 Fabrication flow of multi-layer TFT NAND device

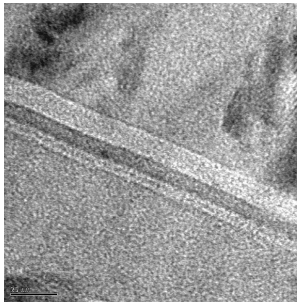


Fig. 1 TEM of the O1/N1/O2/N2/O3 structure for TFT BE-SONOS.

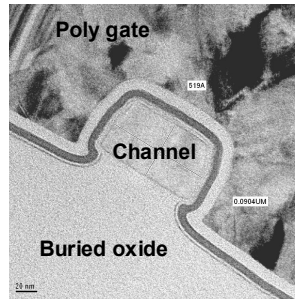


Fig. 2 Tri-gate structure of the TFT device.

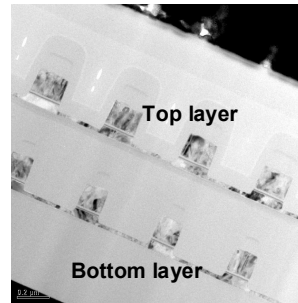


Fig. 3 Channel-length direction of double-layer TFT NAND devices.

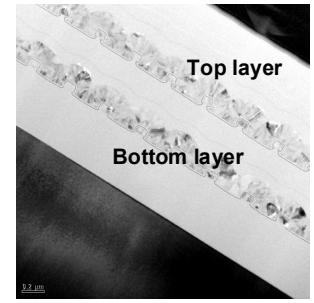


Fig. 4 Channel-width direction of double-layer TFT NAND devices.

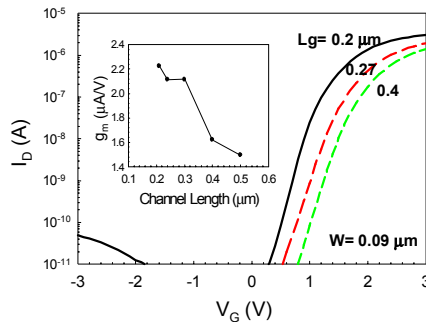


Fig. 5 I_V characteristics for various channel length devices. Inset shows the corresponding transconductance (g_m). The channel width is 0.09 μm .

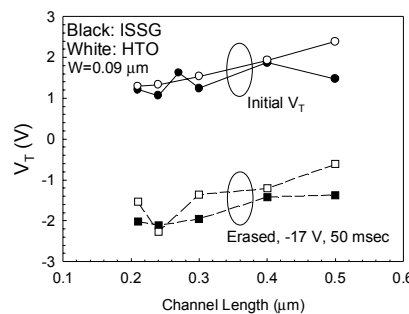


Fig. 6 Short-channel V_T roll-off characteristics for both the initial state and the -FN erased state, with ISSG and HTO top oxide.

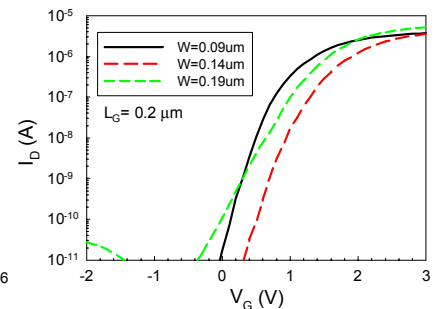


Fig. 7 I_V characteristics for various channel width devices. Wider devices show worse subthreshold behavior. The channel length is 0.2 μm .

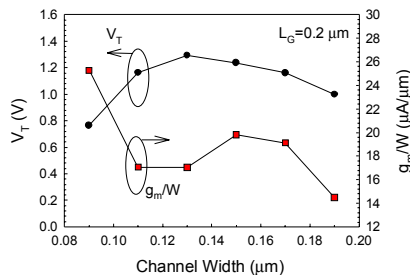


Fig. 8 Narrow-width characteristics. Both normalized transconductance (g_m/W) and V_T are shown. The narrowest device ($W=0.09 \mu\text{m}$) has the best performance.

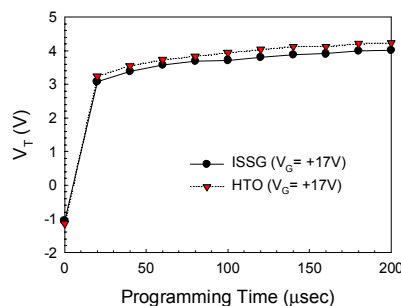


Fig. 9 Comparison of the +FN program characteristics for ISSG and HTO top oxide.

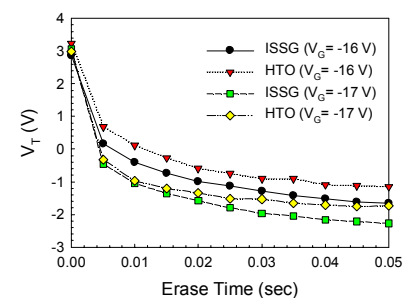


Fig. 10 Comparison of the -FN erase characteristics for ISSG and HTO top oxide.

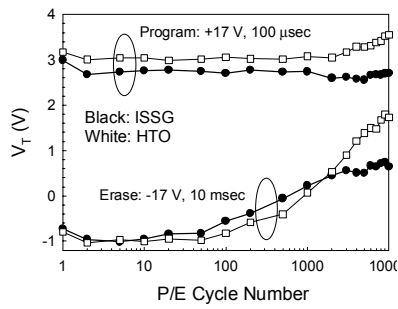


Fig. 11 Dumb-mode P/E cycling endurance (without adjusting the P/E conditions during endurance test). Both ISSG and HTO top oxide are compared.

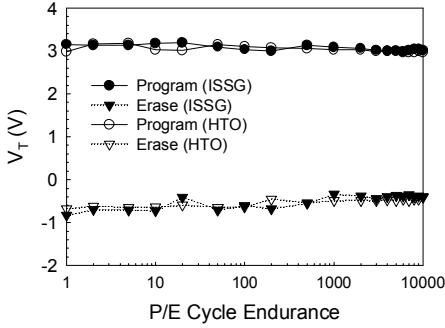


Fig. 12 P/E cycling endurance (with adjusted P/E conditions during endurance test). Both ISSG and HTO top oxide are compared.

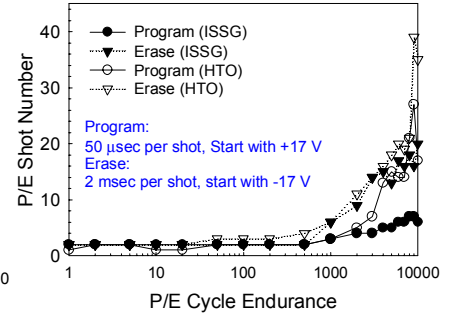


Fig. 13 The required P/E shot number during adjusted P/E tests in Fig. 12. Both ISSG and HTO top oxide are compared.

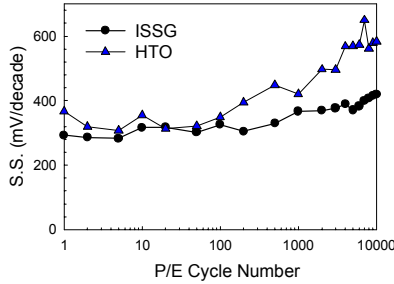


Fig. 14 Subthreshold slope degradation during 10K P/E cycling tests. Both ISSG and HTO top oxide are compared.

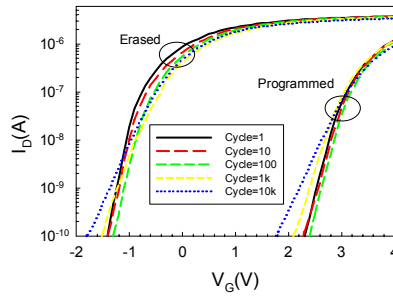


Fig. 15 I_V characteristics of ISSG top oxide device during 10K P/E cycling endurance testing.

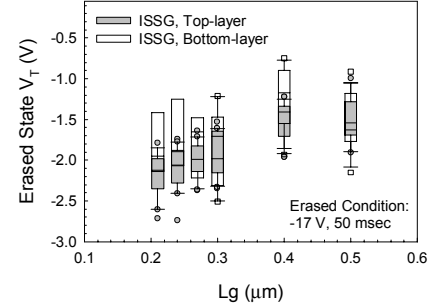


Fig. 16 Short-channel effect comparison of the top and bottom layer TFT NAND devices. A total of 28 identical devices are measured. The devices are in erase saturation state.

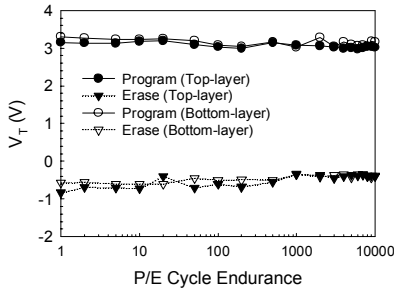


Fig. 17 Adjusted P/E cycling endurance (with adjusted P/E conditions during endurance test). Both bottom and top layers are compared. ISSG top oxide is used.

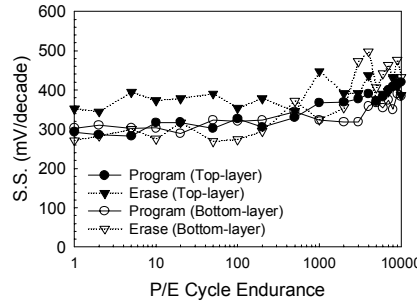


Fig. 18 Subthreshold slope degradation during 10K P/E cycling tests. Both bottom and top layers are compared. ISSG top oxide is used.

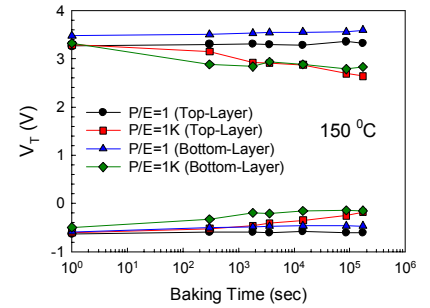


Fig. 19 Retention characteristics. Programmed and erased state with various cycle endurance are compared.

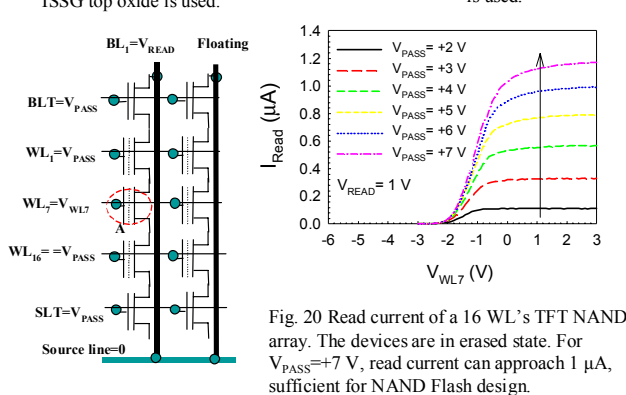


Fig. 20 Read current of a 16 WL's TFT NAND array. The devices are in erased state. For $V_{PASS} = +7$ V, read current can approach 1 μ A, sufficient for NAND Flash design.

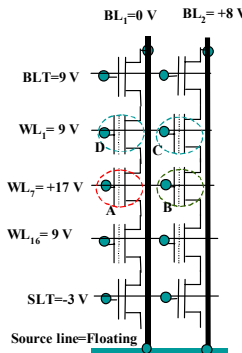


Fig. 21 Program disturb of a TFT NAND array. Cells C and D are under medium field gate disturb, while cell B can be program-inhibited by raising the channel potential.

EXHIBIT C

Experimental Characterization of the Vertical Position of the Trapped Charge in Si Nitride-Based Nonvolatile Memory Cells

Antonio Arreghini, Francesco Driussi, Elisa Vianello, David Esseni, *Senior Member, IEEE*,
Michiel J. van Duuren, Dusan S. Golubović, Nader Akil, and Rob van Schaijk

Abstract—We present a broad set of experiments on silicon nitride-based memories aimed at the investigation of the vertical position of the charge trapped in the nitride layer of silicon-oxide-nitride-oxide-semiconductor (SONOS) memories during program and erase in the tunneling regime. The results obtained for SONOS devices with conventional oxide-nitride-oxide and oxide-nitride-oxide-nitride-oxide gate stacks, as well as with high- κ top dielectric, have been validated by comparing different characterization techniques. It has been shown that, for SONOS cells, the charge centroid is located in the center of the silicon nitride layer, and its position is quite insensitive to the program or erase conditions and to the gate-stack composition.

Index Terms—Carrier separation, charge centroid, silicon nitride, silicon-oxide-nitride-oxide-semiconductor (SONOS).

I. INTRODUCTION

FURTHER scaling of modern floating-gate (FG) memory cells is hampered by the charge loss due to the stress-induced leakage current (SILC) in cycled cells [1]. In particular, SILC prevents the tunnel-oxide thinning below 8 nm [2] and, moreover, raises statistical-reliability issues for the nowadays very large cell arrays [3], [4]. Furthermore, a strong lateral scaling of nonvolatile-memory (NVM) arrays leads to reliability issues related to the parasitic capacitive coupling between adjacent FG cells [5].

Nitride-based memories, like the silicon-oxide-nitride-oxide-semiconductor (SONOS) devices, are good candidates to replace FG cells in the Flash memory chips of the future generations for high-density and embedded applications, owing to the SILC immunity achieved with localized-charge trapping, that guarantee very good endurance characteristics [6], [7]. Moreover, the relatively low program/erase (P/E) voltages and the excellent compatibility with the standard CMOS process of SONOS memories make them suitable for embedded applications [8]. Finally, the absence of erratic tail bits reduces the statistical-reliability issues in very large memory arrays, and the limited capacitive coupling between adjacent SONOS cells

is beneficial for the reliability issues related to the scaling of the spacing between cells.

However, the data retention of SONOS cells programmed and erased by direct tunneling is still a critical issue because of the thin tunnel oxide (≈ 2.5 nm) needed for the tunneling of substrate holes under negative gate voltages [9]. Therefore, in order to optimize the retention characteristics of SONOS memories, an accurate analysis of the trapping properties of the silicon nitride layer that is responsible for the charge storage in the cell is required. In this respect, the knowledge of the location of the charge trapped in the nitride and the type of carriers injected during the P/E operations are not yet complete [10]–[13] despite the importance of these pieces of information in order to understand the charge-loss rates from the nitride toward the gate and substrate and, therefore, to model accurately the retention characteristics.

In this framework, this paper presents an extensive analysis of the vertical position of the charge trapped in the gate dielectric for a wide set of SONOS devices with different oxide-nitride-oxide (ONO) stacks, including samples with nonconventional gate stacks and with a high- κ top dielectric. To accomplish this, we have extensively applied the experimental procedure developed in [14] to extract the trapped-charge centroid, thus fully validating it over several gate stacks for memory cells. Compared with other methods used to probe the charge trapped in the gate stack, our technique is based on the simplest MOSFET electrostatics, and no constraints are imposed to the P/E conditions during its application. In particular, whereas the low-frequency charge pumping (LFCP) [15] can extract the trap profile, but only within a few nanometers inside the ONO stack, our procedure can reach the deeper locations inside the gate stack, thus considering the effect of the whole trapped charge. However, the centroid alone does not provide the detailed knowledge of the charge profile inside the nitride, but it determines the effectiveness of a given distribution of charges in generating a corresponding V_T shift. Therefore, our technique can be seen as complementary to other methods used to probe the trapped-charge profile (e.g., the LFCP), overcoming some of their limitations in terms of accuracy and applicability.

A broad range of P/E conditions, i.e., different programming/erasing gate voltages (V_G), have been tested for each device, and the experimental results for the charge-centroid location have been confirmed by the comparison with those derived with the LFCP method, thus demonstrating the advantages and

Manuscript received November 26, 2007. This work was supported in part by the Italian MIUR under the PRIN 2004 and FIRB RBIP06YSJJ research programs. The review of this paper was arranged by Editor V. R. Rao.

A. Arreghini, F. Driussi, E. Vianello, and D. Esseni are with the Department of Electrical, Mechanical and Management Engineering, University of Udine, 33100 Udine, Italy (e-mail: francesco.driussi@uniud.it).

M. J. van Duuren, D. S. Golubović, N. Akil, and R. van Schaijk are with NXP Semiconductors, 3001 Leuven, Belgium.

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TED.2008.919713

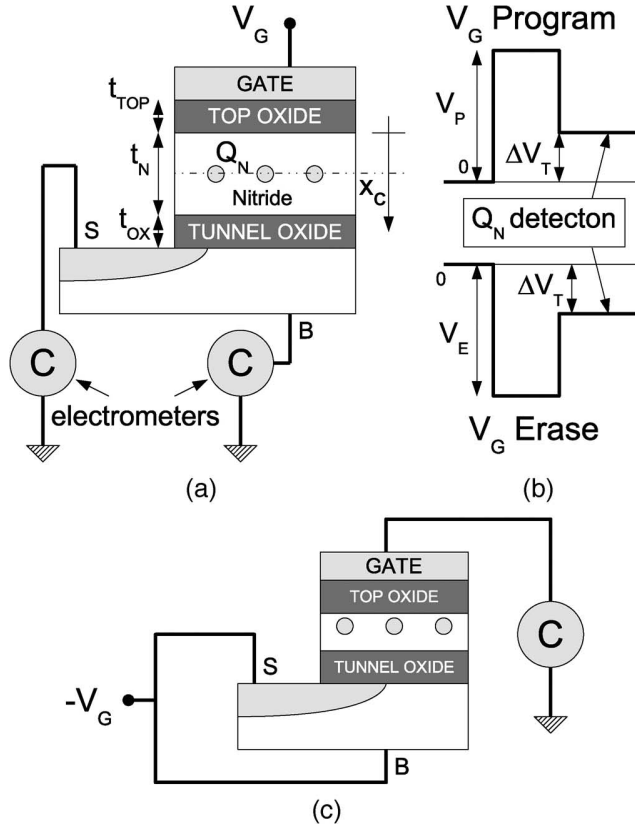


Fig. 1. (a) Measurement setup for carrier separation and charge-centroid extraction. Q_N is the trapped charge, x_C is the centroid position, and t_{TOP} , t_N , and t_{OX} are the thicknesses of the top oxide, nitride layer, and tunnel oxide, respectively. (b) Waveforms for the P/E pulses showing the three-level shape used to obtain a null contribution to the overall current integral of the displacement charge in the semiconductor. (c) Measurement setup for the charge-centroid extraction. By measuring Q_N at the gate while pulsing with a reversed waveform with respect to (b) at the source and bulk terminals, no compensation of possible leakage currents is needed.

limitations of each technique. In a limited number of cases, pMOS SONOS devices were fabricated, and carrier-separation experiments were performed to identify the type of carriers involved in the program/erase operation.

II. EXPERIMENTAL SETUP AND CHARGE-CENTROID-EXTRACTION PROCEDURE

Fig. 1(a) shows the vertical structure of a conventional SONOS device. The charge that is responsible for the threshold-voltage shift (ΔV_T) is supposed to be stored in the silicon nitride layer, which is sandwiched between the tunnel and the top oxides (this assumption will be relaxed later). If we let x_C denote the centroid of the trapped charge measured from the top-oxide/nitride interface [see Fig. 1(a)], then the threshold-voltage shift induced in the cell by the charge Q_N (uniformly injected over the device area A) can be calculated as [14], [16]–[18]

$$\begin{aligned} \Delta V_T &= -\frac{Q_N}{A} \cdot \frac{t_{TOP}}{\epsilon_{OX}} - \frac{1}{\epsilon_N} \int_0^{t_N} x \cdot q(x) dx \\ &= -\frac{Q_N}{A} \left(\frac{t_{TOP}}{\epsilon_{OX}} + \frac{x_C}{\epsilon_N} \right) \end{aligned} \quad (1)$$

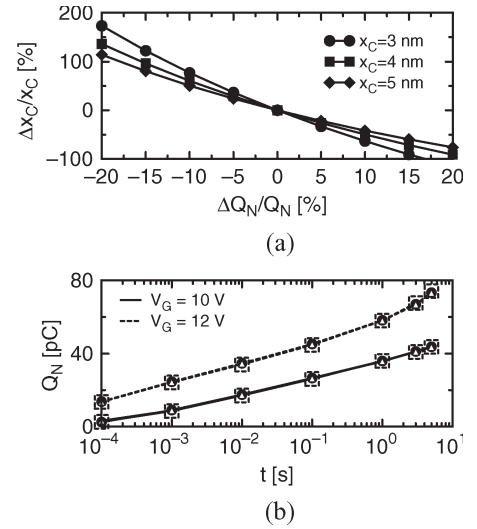


Fig. 2. (a) Sensitivity of the charge centroid on the possible error in the trapped-charge detection. Note that small errors in the measured Q_N lead to large errors in the estimation of x_C . (b) Measured Q_N as a function of the programming pulsewidth for three different devices (circles, squares, and triangles). Note the good repeatability of the measurements obtained with our setup.

where $q(x)$ is the charge density along the nitride depth x , ϵ_{OX} and ϵ_N are the permittivity of SiO_2 and Si_3N_4 , respectively, and t_{TOP} and t_N are the top oxide and nitride thicknesses, respectively. Therefore, if ΔV_T and Q_N are known, the centroid of the charge trapped in the nitride can be univocally determined as

$$x_C = -\epsilon_N \left(\frac{\Delta V_T \cdot A}{Q_N} + \frac{t_{TOP}}{\epsilon_{OX}} \right). \quad (2)$$

Hence, in order to determine x_C by using (2), we have developed a well-controlled experimental setup to directly measure both the programming (erasing) charge Q_N and the corresponding threshold-voltage shift $\Delta V_T = (V_{T \text{ final}} - V_{T \text{ initial}})$ produced by the program (erase) performed in tunneling conditions. For the sake of clarity, it is worth noting that we define program (erase) as the application of gate pulses with positive (negative) V_G , which increase (decrease) the threshold voltage V_T , regardless of the type of the device under test (nMOS or pMOS). In the experimental setup shown in Fig. 1(a), we connected two accurately zeroed electrometers at the source/drain and bulk terminals of the SONOS device. In this way, when an nMOS SONOS cell undergoes a P/E cycle, we can separately measure the contributions to the programming (erasing) charge Q_N of both the electrons detected at source/drain and the holes collected at the bulk, thus performing a carrier-separation experiment.

As for the x_C extraction, an analysis of (2) shows that, while the errors in the detection of ΔV_T are not so critical, x_C is largely sensitive to possible errors on the measured Q_N . With this regard, Fig. 2(a) shows, for a fixed actual centroid position (different symbols), the relative error in the extracted x_C induced by an error in the detection of the charge; we clearly see that even a relatively small error in Q_N can lead to an incorrect estimation of the charge centroid. Therefore, in order to minimize the charge errors, the leakage current through

the source-to-bulk junction has been compensated for, first, by measuring its spurious contribution to the charge during a pulse train with a reduced V_G value that prevents the injection into the gate stack and, then, by subtracting the obtained value from the Q_N measured during the P/E pulse with large V_G .

Alternatively, because the capacitor is a two-terminal device, the injected charge can be measured also by reversing the roles of the device contacts, hence, through an electrometer connected to the gate, while pulsing on the bulk and source terminals [Fig. 1(c); source and bulk are short-circuited]. In this configuration, a $-V_G$ value is applied during the pulse. In such a way, although carrier separation is not possible, no leakage compensation is needed on the gate terminal (where the charge is measured). Therefore, for the purpose of extracting the charge centroid, the measurement of Q_N by means of the experimental setup in Fig. 1(c) is highly preferable.

Moreover, in order to avoid the spurious contribution to Q_N due to a net, which is the nonnull-time integral of the displacement current, we forced the same $(V_G - V_T)$, hence the same electric field E_{OX} at the Si/SiO₂ interface, which is before and after the program/erase pulse. This was obtained by applying to the gate three-level pulses, as shown in Fig. 1(b), where the final base level of the pulse is displaced by ΔV_T with respect to the initial base level. This approach requires the knowledge in advance of the ΔV_T induced by each program (erase) pulse, and therefore, the P/E characteristics have been preliminarily measured. Afterwards, the measured ΔV_T 's have been used to set the pulse shape and measure the Q_N , as shown in Fig. 1(c). Because the programming and the erasing are performed by applying a sequence of several P/E pulses, for each measurement, we monitored both the cumulative charge Q_N involved in the whole P/E sequence and the partial charge q_N injected during each individual P/E pulse.

As shown in Fig. 2(b), by using this measurement scheme on a given SONOS device, we could achieve a very good reproducibility of the measured Q_N and ΔV_T among different runs and dies, provided that all the P/E characteristics start from the same initial V_T value. An *ad hoc* reset procedure of the device consisting of several mild P/E cycles allowed us to enforce always the same starting V_T value within tens of millivolts. The possible degradation produced by the P/E cycling has been monitored by means of charge-pumping measurements, and it was found to be negligible in the range of relatively few P/E cycles used for our characterization experiments.

III. DEVICE DESCRIPTION

The devices used in this paper are nMOS and pMOS SONOS capacitors with a surrounding source diffusion and a gate area of $A = 3100 \mu\text{m}^2$. For both nMOS and pMOS devices, n+ and p+ polygates are available, and the substrate doping is approximately 10^{17} cm^{-3} . The vertical dimensions of the ONO gate stack have been extracted by means of transmission electron microscopy (TEM), ellipsometry, and capacitance–voltage (C – V) measurements. The nominal thickness of the thermal tunnel oxide is $t_{OX} = 2.2 \text{ nm}$ for all SONOS cells, whereas two thicknesses are available for both the dichlorosilane Si₃N₄ layer ($t_N = 3.3$ and 5.7 nm , respectively) and the high-temperature-

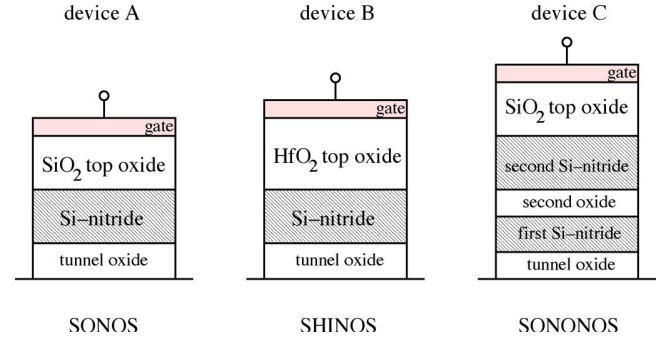


Fig. 3. Gate-stack composition of the devices used in this paper.

oxide (HTO) top oxide [$t_{TOP} = 8.9$ and 5.5 nm , respectively; Fig. 3, device A].

A gate stack featuring a 11.8-nm-thick HfO₂ layer as top dielectric, a 5.7-nm-thick Si nitride, and a 2.2-nm-thick tunnel oxide was also available (hereafter denoted as SHINOS; Fig. 3, device B).

Finally, the nonconventional gate-stack structure shown in Fig. 3 (device C, hereafter denoted as SONONOS) has been also used for the experiments related to the charge-centroid determination. This device is composed of two nitride layers sandwiched between the tunnel and top oxides and separated by an additional SiO₂ layer (second oxide) formed by the reoxidation of part of the first silicon nitride layer (the one on top of the tunnel oxide). In this device, the thicknesses extracted by means of TEM and C – V measurements are 1.3, 2.8, 1.5, 5.0, and 7.2 nm for the tunnel oxide, the first nitride, the second oxide, the second nitride, and the top oxide, respectively (Fig. 3, device C). All the measured devices are free of the passivation issue discussed in [19].

IV. CARRIER-SEPARATION EXPERIMENTS

Carrier-separation experiments are suitable to investigate quantitatively the contribution of electrons and holes during the program and erase operations. Previous experiments showed that, during program, the charge Q_N is almost entirely due to electrons [11], [13], [14], [17], whereas during erase, the contribution of holes can be significant [9], [14]. In this regard, we first successfully verified with the experimental setup in Fig. 1(a) that Q_N is due to electrons during program [14]. In analogy with the case of quantum-yield experiments, the hole recombination with electrons in the inversion layer is a very unlikely process [20], [21], and, thus, we conclude in agreement with Bachhofer *et al.* [11], Wang and White [13], and Libsch and White [17] that the hole detrapping is negligible during this phase.

In the case of erasing, because our nMOS SONOS test structures are rather large ($20 \times 155 \mu\text{m}$), the electrons that are possibly detrapped from the nitride must diffuse for several micrometers in the substrate before they can reach the source contact; thus, the recombination with substrate holes can take place, and the carrier-separation technique may underestimate the contribution to the erase of the detrapped electrons. The experiments on nMOS devices revealed that the measured Q_N during erase is mainly due to holes [14]; hence, either

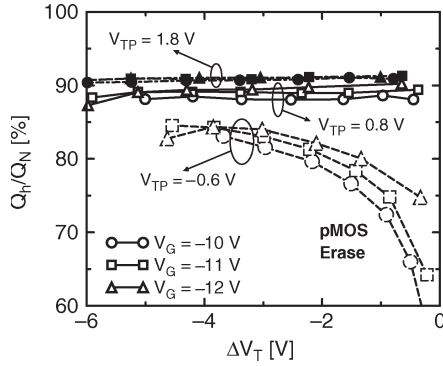


Fig. 4. Fraction of the charge due to holes measured during the pMOS SONOS erase as a function of ΔV_T for different values of V_G and of the initial threshold voltage V_{TP} . Note the decreasing contribution of holes for decreasing V_{TP} . $Q_h, Q_e > 0$ during erase.

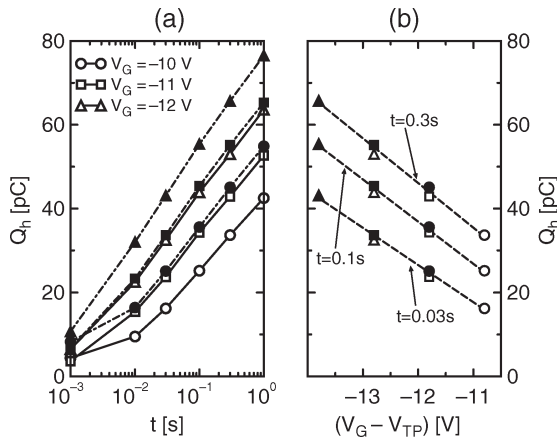


Fig. 5. Hole contribution to pMOS SONOS erase as a function of the (a) erasing time and (b) $(V_G - V_{TP})$. Open symbols for $V_{TP} = 0.8$ V and filled symbols for $V_{TP} = 1.8$ V.

the electron detrapping is very small or it is masked by the recombination in the p -substrate.

The accuracy of the carrier-separation experiments for determining the contribution of electrons and holes (Q_e and Q_h , respectively) to the erase is significantly improved if we use pMOS SONOS capacitors. In fact, under negative V_G , the holes are supplied by the inversion layer and are measured at the source, whereas the detrapped electrons go through the very thin hole inversion layer and, then, are collected at the bulk terminal of the n -type substrate [20], [21] so that the probability of recombination with holes is drastically reduced with respect to the experiment of the erase in nMOS SONOS.

Fig. 4 shows the carrier-separation measurements performed during the erase transient in pMOS SONOS devices, which, according to our definition, correspond to a reduction of the threshold voltage, i.e., to negative ΔV_T values. We see how the hole contribution Q_h to Q_N depends on the initial programmed state V_{TP} . In particular, for $V_{TP} = -0.6$ V, we observe a reduced contribution of holes, hence a large contribution of electrons detrapped from the nitride layer and a strong ΔV_T dependence of the (Q_h/Q_N) ratio. At higher V_{TP} , the relative contribution of holes remains constant versus ΔV_T , regardless of the erasing gate voltage V_G . Moreover, if we replot Q_h as in Fig. 5, we see how the hole component directly depends on the

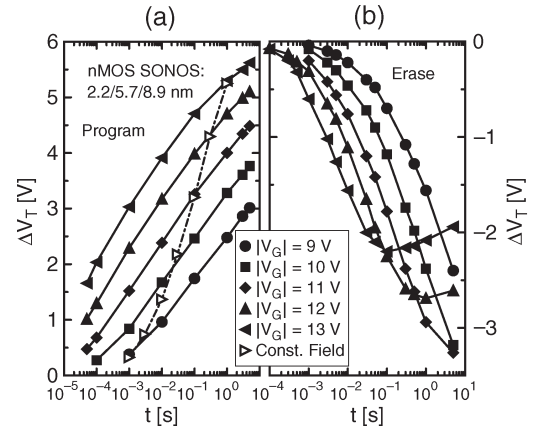


Fig. 6. (a) Program and (b) erase characteristics of an nMOS SONOS capacitor at constant gate voltage. Open triangles correspond to programming at constant tunnel-oxide field, which is obtained by compensating at each program pulse the ΔV_T produced by the previous pulse.

initial gate overdrive voltage ($V_G - V_{TP}$). This latter result can be justified by considering that both the tunneling rate and the hole-inversion-layer density depend on the $(V_G - V_{TP})$ value.

V. POSITION OF THE CHARGE CENTROID

After the identification of the different roles played by electrons and holes during P/E, we focused our attention on the extraction of the charge centroid x_C by using (2). In particular, to study the x_C dependence on the P/E conditions and the device structure, the experimental procedure described in Section II has been used to measure Q_N and the induced ΔV_T for several V_G values and for devices featuring different ONO gate stacks.

A. Experimental Results for SONOS and SHINOS Devices

Fig. 6 shows the P/E characteristics as functions of the total program (erase) time t of the nMOS SONOS with $t_{OX} = 2.2$ nm, $t_N = 5.7$ nm, and $t_{TOP} = 8.9$ nm (2.2/5.7/8.9). The ΔV_T 's have been evaluated by measuring the shifts of the $C-V$ curve after each P/E pulse. Then, the capacitor was reset to the initial state (by mild P/E cycling), and the ΔV_T data have been used to set the shape of the pulses [see Fig. 1(b)] during which we directly measured the injected Q_N with the experimental setup in Fig. 1(c). Finally, in Fig. 7(a), we have calculated the x_C of the cumulative charge Q_N . Each point of the curve is the centroid of the total charge Q_N injected from the beginning of the P/E characteristic to the time corresponding to the ΔV_T value reported in the x -axis in Fig. 7(a).

We observe that the centroids for program and erase are very consistent with each other; at the beginning of the program (erase) phase, x_C is located near the middle of the nitride layer, which is in agreement with the results for HTO top oxides reported in [22]. Then, x_C slowly moves toward the top oxide. Moreover, the data in Fig. 7(a) do not show any dependence on the applied gate voltage V_G . In this respect, in order to further investigate a possible bias dependence of x_C , we have also performed measurements by keeping E_{OX}

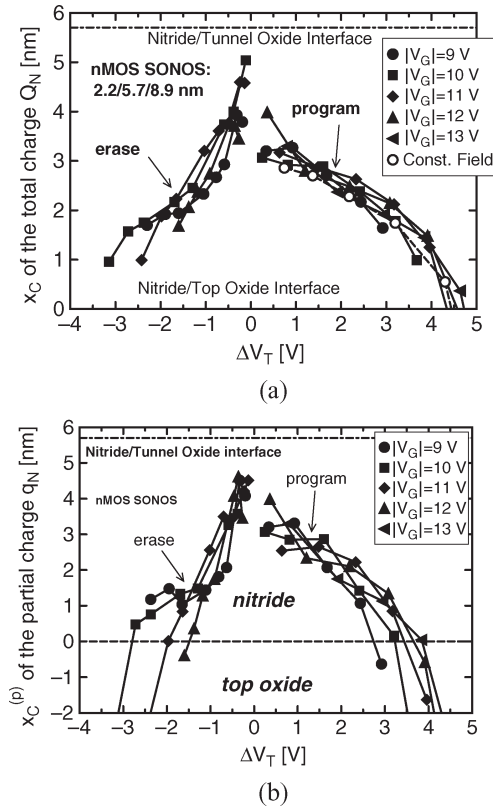


Fig. 7. (a) Centroid of the cumulative charge trapped in the nitride of the nMOS SONOS capacitor (2.2/5.7/8.9-nm ONO stack) extracted for the program and erase pulses at different V_G 's. Open circles correspond to the programming at constant electric field across the tunnel oxide (V_G is adjusted after each subsequent pulse). (b) Centroid of the partial charge injected into the nitride during each pulse of the programming/erasing of the nMOS SONOS (2.2/5.7/8.9-nm ONO stack). Note how during the last program/erase pulses (large $|\Delta V_T|$) the centroid of the partial charge is located in the top oxide, indicating that part of the injected charge has been lost through the gate.

at the Si/SiO₂ interface approximately constant during program [Figs. 6(a) and 7(a); open symbols] by increasing V_G after each pulse for an amount equal to the corresponding ΔV_T . This, of course, leads to a faster programming [Fig. 6(a); open symbols], but x_C still agrees well with the data obtained at constant V_G [Fig. 7(a)]. Therefore, we can conclude that there is no dependence of x_C on the injection conditions, at least in the bias range explored in this paper.

For $\Delta V_T > 4$ V, the x_C values move toward the top oxide and eventually become negative. The negative x_C , however, is most likely an artifact due to an electron leakage through the top oxide at the largest programming times, hence largest ΔV_T 's. In fact, when the cell is strongly programmed and the nitride is almost completely charged, the electric field across the top oxide is maximum, and the electrons injected in the nitride can tunnel out of the storage layer during the last part of the programming phase. This charge leakage offsets our estimate of Q_N , which hence appears less effective with respect to the measured threshold-voltage shift and results in a smaller value of x_C [see (2)]. In order to more precisely assess the ΔV_T range where the determination of x_C is reliable, in Fig. 7(b), we calculated the centroid $x_C^{(p)}$ of the partial charge packet q_N injected during the last pulse of the P/E sequence used to reach

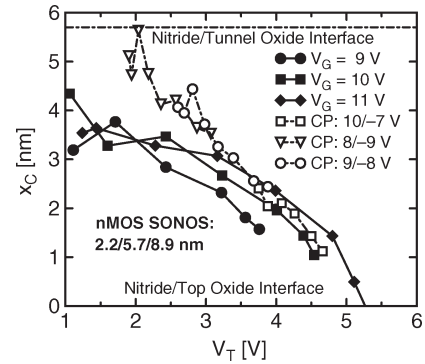


Fig. 8. Centroid x_C for the nMOS SONOS capacitor (2.2/5.7/8.9-nm ONO stack) extracted during programming. The data obtained with the (filled symbols) direct-measurement technique summarized by (2) are in good agreement with those provided by the (open symbols) LFCP experiments.

a given ΔV_T value, hence injected between two adjacent points of the curves in Fig. 7(a). The centroid $x_C^{(p)}$ is calculated by inserting in (2) the charge q_N and the threshold-voltage shift corresponding to the single (individual) P/E pulse. It is worth to note how, for $\Delta V_T > 3$ V, $x_C^{(p)}$ is negative, thus indicating that we are measuring a charge packet q_N that is larger than the one needed to justify the partial threshold-voltage shift obtained during the last P/E pulse. Hence, part of the charge q_N must have been lost through the gate. For $\Delta V_T < 3$ V, on the other hand, the $x_C^{(p)}$ of each partial charge packet is located well inside the nitride, thus indicating that we are not losing charge through the gate and validating the x_C values extracted for the cumulative charge Q_N in this range of ΔV_T values.

Moreover, the analysis of the $x_C^{(p)}$ values shows that the centroid position is independent of the starting V_T , because each point in Fig. 7(b) corresponds to the charge centroid obtained, starting from a different V_T level.

During erase, $x_C^{(p)}$ becomes negative when the electron injection from the gate takes place. In this situation, a steady-state current flows from the substrate to the gate, thus yielding a saturation of the erase characteristics that limit the observable ΔV_T range [Fig. 6(b)]. The charge measured in this condition is largely exceeding the charge responsible for the threshold-voltage shift, and therefore, the extracted x_C becomes unreliable while approaching this region. On the basis of such observations, we estimated that, for the measured device, the x_C values are reliable for $-2 < \Delta V_T < 3$ V and, in this ΔV_T range, the charge centroid stays close to the center of the nitride layer and it only slightly moves toward the top oxide for both program and erase [14], [18], [22].

As stated in Section IV, the carrier-separation experiments showed that, in our nMOS SONOS capacitors, the measured Q_N during program is essentially due to electrons whereas, during erase, it is mainly due to holes. Thus, the net result of a complete P/E cycle is a recombination charge, and we can also use an LFCP technique to detect such a recombined charge Q_N [15]. Fig. 8 shows the x_C values calculated by using Q_N detected by means of LFCP with different base and top pulse levels. We see that the results agree well with the x_C obtained with the direct measure of Q_N on the same device, except at small ΔV_T , where the LFCP method underestimates Q_N , as

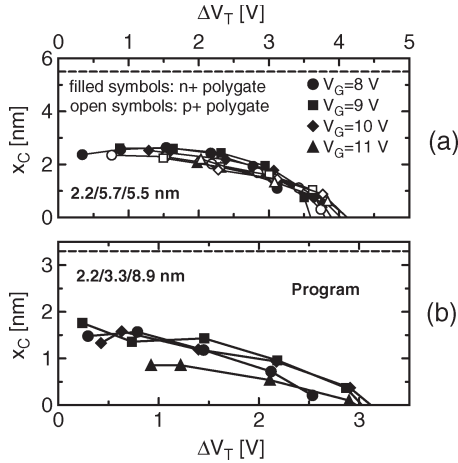


Fig. 9. Centroid x_C extracted during the programming of SONOS devices, featuring two different ONO stacks: (a) 2.2/5.7/5.5 and (b) 2.2/3.3/8.9. Filled and open symbols are the data for SONOS with n+ polygate and p+ polygate, respectively. The dashed line indicates the tunnel-oxide/silicon-nitride interface.

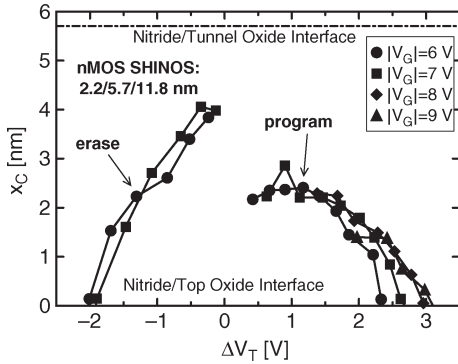


Fig. 10. Charge centroid extracted during the program and the erase at different V_G 's for the nMOS SHINOS cells, featuring a HfO_2 top oxide of 11.8 nm.

shown in [14], thus resulting in a deviation of x_C toward the tunnel oxide [see (2)].

Fig. 9 shows the centroid x_C extracted during the programming of SONOS capacitors with different top oxide [$t_{\text{TOP}} = 5.5$ nm; (a)] and silicon nitride thicknesses [$t_N = 3.3$ nm; (b)], whereas all other stack dimensions are the same, as shown in Fig. 7. Also, in this case, we find that x_C is located near the middle of the nitride (slowly shifting toward the top oxide as programming continues) and no gate bias dependence is observed. Moreover, there is no significant dependence of the measured x_C on the polygate type (open and filled symbols). Fig. 10 shows the x_C values extracted from the SHINOS cells featuring the HfO_2 top dielectric (Fig. 3, device B). Also, in this case, x_C is located inside the nitride, no dependence on V_G is observed, and the data for program and erase are consistent with each other. The results obtained for the pMOS SONOS devices are very similar [14].

As a summary, Fig. 11 shows the centroid position normalized to the nitride thickness t_N for all the measured SONOS devices. We see how, for moderate ΔV_T values, x_C is always located near the center of the nitride irrespective of the gate-stack composition and the substrate doping type.

Authorized licensed use limited to: IEEE - Staff. Downloaded on August 05, 2025 at 18:39:18 UTC from IEEE Xplore. Restrictions apply.

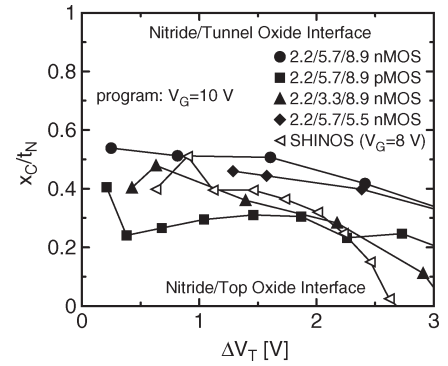


Fig. 11. Normalized position along the nitride layer of the charge centroid extracted from different devices. Note that the relative position of x_C is near the center of nitride for all the samples, as long as ΔV_T is not very large.

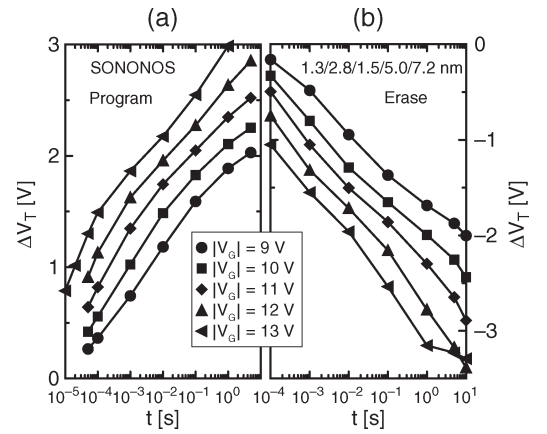


Fig. 12. (a) Program and (b) erase characteristics of the SONONOS cell for different V_G values.

B. Experimental Results for SONONOS Devices

We applied the charge-centroid-extraction procedure also to the SONONOS devices featuring the gate-stack structure shown in Fig. 3 (device C). Fig. 12 shows the P/E characteristics of this SONONOS capacitor, which show a faster programming and erasing compared with SONOS cells (see Fig. 6). In order to perform the x_C extraction accurately, particular care has been taken in the ΔV_T measurement. In fact, the application of even small V_G values produces appreciable threshold-voltage shifts, due to the very thin tunnel oxide (1.3 nm). This results in a considerable hysteresis in the $C-V$ characteristics obtained, spanning a relatively large V_G range [Fig. 13(a)]. In our capacitors, the ΔV_T detection is performed by measuring the shift of the $C-V$ characteristic, and this can be accomplished by detecting the capacitance at different points, hence V_G values. However, in the SONONOS device, we had to choose an appropriate evaluation point for the ΔV_T [Fig. 13(a), point B] in order to avoid the application of too large $(V_G - V_T)$ values that induce a spurious programming during the ΔV_T extraction, as in the case of the choice of the point A in the $C-V$ curve in Fig. 13(a). In the latter case, the gate disturbance leads to a parasitic threshold-voltage shift that depends on the measurement time [Fig. 13(b), in squares]. The experimental results presented in the following have been extracted by detecting the ΔV_T at

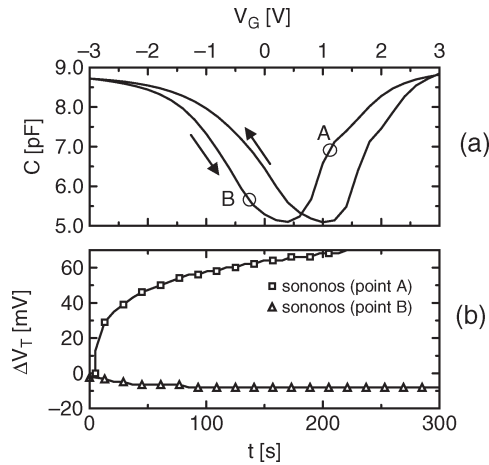


Fig. 13. (a) C - V curves of the SONONOS, showing a large hysteresis during the measurement. The arrows indicate the direction of the V_G sweep. (b) The ΔV_T value is measured as a shift of the C - V curve. The selection of the right evaluation point in the C - V curve avoids a spurious V_T shift induced by a gate disturb during the detection of the ΔV_T .

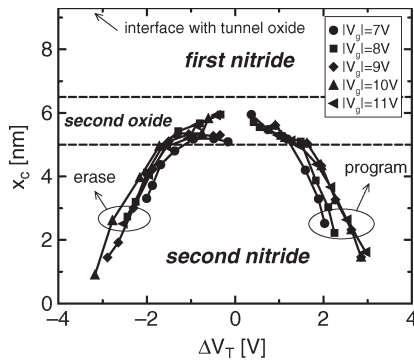


Fig. 14. Charge centroid extracted from the SONONOS devices. Initially, the charge is trapped near the interface between the first nitride and the second SiO_2 layer, and then, it fills the traps in the center of the second nitride. See Fig. 3 (device C) for the gate-stack definition.

(triangles) point B, where the gate disturbance is essentially negligible.

On the other hand, it is worth noting that, in the SONOS samples, the discussed issue is not present due to the fact that the C - V curve is not hysteretic because of the thicker tunnel oxide. In this case, the ΔV_T detection can be done at any point of the C - V curve, because the application of the V_G bias that is necessary for the capacitance measurement does not also produce spurious threshold-voltage shifts if this V_G value is applied to the device for long times.

The x_C values extracted from the SONONOS cells are shown in Fig. 14. Also, in this case, there is no dependence on the applied P/E conditions, and the data for program and erase are consistent. However, unlike standard SONOS devices, SONONOS structures exhibit a centroid x_C that is initially located very close to the second oxide layer (Fig. 3, device C). This is probably due to the fact that the second oxide has been formed by the reoxidation of part of the first nitride layer. This oxide-fabrication method is known to produce a large number of traps at the interface between the first silicon nitride layer and the second oxide layer [22]–[24]. The density of these interface traps can be larger than the concentration in the bulk

of the nitride layer, and the results in Fig. 14 are consistent with the assumption that the carriers are mainly trapped at the $\text{Si}_3\text{N}_4/\text{SiO}_2$ interface at the beginning of the program/erase operation, namely, for small $|\Delta V_T|$ values. Because in this case, x_C is located in the SiO_2 layer that has a different permittivity ϵ_{OX} with respect to nitride, for the evaluation of the x_C values, (2) has been modified as follows:

$$x_C = -\epsilon_{\text{OX}} \left(\frac{\Delta V_T A}{Q_N} + \frac{t_N}{\epsilon_N} \right) + t_N - t_{\text{TOP}}. \quad (3)$$

By increasing the program (erase) level, electrons (holes) start filling the bulk traps located in the second nitride film, and x_C moves to the middle of the nitride thickness as in the case of SONOS structures. It seems, however, that no significant amount of charge is trapped in the first nitride layer, although a comparable concentration of free traps should be available with respect to the second nitride layer.

These results also indicate that our experimental technique is indeed sensitive; in fact, it can detect changes of the x_C value due to different trapping properties in the gate stack.

VI. CONCLUSION

We have extensively used carrier-separation experiments with direct measurement of the charge injected during the program and erase operations of SONOS memory cells to study the type of carrier that is mainly involved in each operation as well as the location of the trapped charge Q_N . Preliminary experiments showed that, whereas Q_N is entirely due to electrons during program, during erase, there is a significant contribution of holes, which follows a $(V_G - V_{\text{TP}})$ dependence.

Concerning the extraction of the charge centroid, a wide set of experiments in several P/E conditions on SONOS cells with different gate-stack structures has consistently shown that the centroid of the cumulative charge (x_C) is always located close to the center of the trapping layer and it slowly shifts toward the top oxide irrespective of the applied P/E voltages. Moreover, the x_C values do not significantly depend on the top oxide and nitride thicknesses, the composition of the top dielectric, or the substrate and gate doping type.

The centroid results for the SONONOS gate stack instead show that the carriers are initially trapped near the intermediate oxide layer likely because the reoxidation of the first nitride layer has generated a large number of traps at the $\text{Si}_3\text{N}_4/\text{SiO}_2$ interface [22]–[24]. For larger P/E levels, the carriers start filling the bulk traps of the second Si_3N_4 layer, so that the centroid moves toward the middle of the second nitride layer, which is consistent with the results obtained for SONOS devices. Also, for these devices, the behavior of x_C is essentially independent of the P/E conditions.

This large set of experimental data demonstrates the sensitivity of the characterization procedure proposed in this paper, which is thus useful to study the trapping properties of different gate-stack structures and support the design of the nitride-based NVM cells. The data also represent a valid reference for the calibration of program/erase and retention models of Si nitride-based NVM [25].

ACKNOWLEDGMENT

The authors would like to thank Prof. L. Selmi for the fruitful discussions and the constant encouragement. Parts of the samples were fabricated in the framework of the MEDEA+ project NEMeSyS.

REFERENCES

- [1] J. Blauwe, J. Van Houdt, D. Wellekens, R. Degraeve, P. Roussel, L. Haspeslagh, L. Deferm, G. Groeseneken, and H. E. Maes, "A new quantitative model to predict SILC-related disturb characteristics in flash E²PROM devices," in *IEDM Tech. Dig.*, 1996, p. 343.
- [2] K. Yoshikawa, S. Mori, E. Sakagami, N. Arai, Y. Kaneko, and Y. Ohshima, "A flash EEPROM cell scaling including tunnel oxide limitations," in *Proc. Eur. Solid State Device Res. Conf.*, 1990, pp. 169–172.
- [3] F. Driussi, F. Widdershoven, D. Esseni, L. Selmi, and M. van Duuren, "Experimental characterization of statistically independent defects in gate dielectrics—Part I: Description and validation of the model," *IEEE Trans. Electron Devices*, vol. 52, no. 5, pp. 942–948, May 2005.
- [4] D. Ielmini, A. S. Spinelli, A. L. Lacaita, and M. Duuren, "Correlated defect generation in thin oxides and its impact on Flash reliability," in *IEDM Tech. Dig.*, 2002, pp. 143–146.
- [5] A. Ghetti, L. Bortesi, and L. Vendrame, "3D Simulation study of gate coupling and gate cross-interference in advanced floating gate non-volatile memories," *Solid State Electron.*, vol. 49, no. 11, pp. 1805–1812, Nov. 2005.
- [6] M. H. White, D. A. Adams, and J. Bu, "On the go with SONOS," *IEEE Circuits Devices Mag.*, vol. 16, no. 4, pp. 22–31, Jul. 2000.
- [7] R. van Schaijk, M. van Duuren, P. Goarin, W. Y. Mei, and K. van der Jeugd, "Reliability of embedded SONOS memories," in *Proc. ESSDERC*, 2004, pp. 159–162.
- [8] Y. Kamigaki and S. Minami, "MNOS nonvolatile semiconductor memory technology: Present and future," *IEICE Trans. Electron.*, vol. E84-C, no. 6, pp. 713–723, Jun. 2001.
- [9] E. Suzuki, K. Miura, Y. R. Tsay, and D. Schroder, "Hole and electron current transport in metal-oxide-nitride-oxide-silicon memory structures," *IEEE Trans. Electron Devices*, vol. 36, no. 6, pp. 1145–1149, Jun. 1989.
- [10] Y. Yang, A. Purwar, and M. White, "Reliability considerations in scaled SONOS nonvolatile memory devices," *Solid State Electron.*, vol. 43, no. 11, pp. 2025–2032, Nov. 1999.
- [11] H. Bachhofer, H. Reisinger, E. Bertagnolli, and H. Philipsborn, "Transient conduction in multilayered silicon-oxide-nitride-oxide semiconductor structures," *J. Appl. Phys.*, vol. 89, no. 5, pp. 2791–2800, Mar. 2001.
- [12] S. Chung, P.-Y. Chiang, G. Chou, C.-T. Huang, P. Chen, C.-H. Chu, and C.-H. Hsu, "A novel leakage current separation technique in a direct tunneling regime gate oxide SONOS memory cell," in *IEDM Tech. Dig.*, 2003, p. 617.
- [13] Y. Wang and M. White, "An analytical retention model for SONOS nonvolatile memory devices in the excess electron state," *Solid State Electron.*, vol. 49, no. 1, pp. 97–107, Jan. 2005.
- [14] A. Arreghini, F. Driussi, D. Esseni, L. Selmi, M. van Duuren, and R. van Schaijk, "Experimental extraction of the charge centroid and of the charge type in the P/E operation of SONOS memory cells," in *IEDM Tech. Dig.*, 2006, pp. 499–502.
- [15] A. Arreghini, F. Driussi, D. Esseni, L. Selmi, M. van Duuren, and R. van Schaijk, "New charge pumping model for the analysis of the spatial trap distribution in the nitride layer of SONOS devices," *Microelectron. Eng.*, vol. 80, pp. 333–336, 2005.
- [16] S. M. Sze, *Physics of Semiconductor Devices*. Hoboken, NJ: Wiley, 1981.
- [17] F. R. Libsch and M. White, "Charge transport and storage of low programming voltage SONOS/MONOS memory devices," *Solid State Electron.*, vol. 33, no. 1, pp. 105–126, Jan. 1990.
- [18] P.-Y. Du, H.-T. Lue, S.-Y. Wang, E.-K. Lai, T.-Y. Huang, K.-Y. Hsieh, R. Liu, and C.-Y. Lu, "Study of the gate-sensing and channel-sensing transient analysis method for monitoring the charge vertical location of SONOS-type devices," *IEEE Trans. Device Mater. Rel.*, vol. 7, no. 3, pp. 407–419, Sep. 2007.
- [19] F. Driussi, L. Selmi, N. Akil, M. van Duuren, and R. van Schaijk, "On the passivation of interface states in SONOS test structures: Impact of device layout and annealing process," in *Proc. IEEE ICMTS*, 2006, pp. 51–55.
- [20] C. Chang, C. Hu, and R. Brodersen, "Quantum yield of electron impact ionization in silicon," *J. Appl. Phys.*, vol. 57, no. 2, pp. 302–309, Jan. 1985.
- [21] A. Ghetti, M. Alam, J. Bude, D. Monroe, E. Sangiorgi, and H. Vaidya, "Stress induced leakage current analysis via quantum yield experiments," *IEEE Trans. Electron Devices*, vol. 47, no. 7, pp. 1341–1348, Jul. 2000.
- [22] H.-T. Lue, Y.-H. Shih, K.-Y. Hsieh, R. Liu, and C.-Y. Lu, "A transient analysis method to characterize the trap vertical location in nitride-trapping devices," *IEEE Electron Device Lett.*, vol. 25, no. 12, pp. 816–818, Dec. 2004.
- [23] E. Suzuki, H. Hiraishi, K. Ishii, and Y. Hayashi, "A low-voltage alterable EEPROM with metal-oxide-nitride-oxide-semiconductor (MONOS) structures," *IEEE Trans. Electron Devices*, vol. ED-30, no. 2, pp. 122–127, Feb. 1983.
- [24] M. Naich, G. Rosenman, and Y. Roizin, "Profiling of deep traps in silicon oxide-nitride-oxide structures," *Thin Solid Films*, vol. 417, no. 1/2, pp. 166–169, Jan. 2005.
- [25] A. Arreghini, N. Akil, F. Driussi, D. Esseni, L. Selmi, and M. van Duuren, "Characterization and modeling of long term retention in SONOS non volatile memories," in *Proc. ESSDERC*, 2007, pp. 406–409.



Antonio Arreghini was born in Udine, Italy, in 1979. He received the Master Laurea degree in electronic engineering from the University of Udine, Udine, Italy, in 2004, where he is currently working toward the Ph.D. degree in industrial and information engineering in the Department of Electrical, Mechanical and Management Engineering.

His research topics include the modeling and the experimental characterization of silicon nitride-based nonvolatile memory cells. From July 2006 to July 2007, he was a Visiting Scientist with NXP Semiconductors, Leuven, Belgium, where he worked on modeling and simulations of long-retention program and erase transients of silicon nitride-based nonvolatile memories.



Francesco Driussi received the Laurea and Ph.D. degrees in electronic engineering from the University of Udine, Udine, Italy, in 2000 and 2004, respectively, working on the reliability of Flash memory devices.

From October 2002 to September 2003, he held a student position with Philips Research Leuven, Leuven, Belgium. Since 2005, he became a Research Associate with the University of Udine, where he is currently with the Department of Electrical, Mechanical and Management Engineering. His research

activities are mainly in the field of nonvolatile memory cell reliability and, in particular, the characterization of device degradation and gate dielectric reliability. In particular, he has worked on substrate-enhanced hot-electron phenomena, and he has investigated on their implications for Flash electrically erasable programmable read-only memory devices. More recently, he worked on the development of physical and statistical models for the study of stress-induced leakage current in large Flash memory arrays and of the oxide trap generation and distribution. He is also working on the characterization and modeling of SONOS memory cells and the trapping properties of silicon nitride. Part of his current activities are also in the field of the experimental characterization and modeling of the carrier mobility in MOSFET devices, featuring strained silicon and high- κ gate stacks.



Elisa Vianello was born in Rovereto, Italy, in 1981. She received the Laurea degree in electronic engineering from the University of Udine, Udine, Italy, in 2006, where she is currently working toward the Ph.D. degree in electronic engineering in the Department of Electrical, Mechanical and Management Engineering.

Her research interests are in the field of the characterization and modeling of nonvolatile memory cells and, in particular, SONOS memory cells.



David Esseni (S'98–M'00–SM'06) received the Laurea and Ph.D. degrees in electronic engineering from the University of Bologna, Bologna, Italy, in 1994 and 1998, respectively.

In 1999, he became an Assistant Professor and, later on, an Associate Professor with the University of Udine, Udine, Italy, where he is currently in the Department of Electrical, Mechanical and Management Engineering. In 2000, he was a Visiting Scientist with Bell Laboratories, Lucent Technologies, Murray Hill, NJ. His research activities include characterization techniques for electrical parameters in MOS devices and several aspects related to the hot-electron phenomena in MOSFETs and Flash electrically erasable programmable read-only memories (EEPROMs). In particular, he has studied the reliability of thin oxide transistors, the statistical distribution of stress-induced leakage current in nonvolatile-memory arrays, and the low-voltage and substrate-enhanced hot-electron injection for applications to Flash EEPROM devices. More recently, he has been involved in the experimental characterization and modeling of low-field mobility in ultrathin silicon-on-insulator MOSFETs and in MOSFETs with ultrathin gate oxides. His modeling interests also include the quantization in MOS devices beyond the effective mass approximation and the semiclassical Monte Carlo approach for decananometric MOSFETs.

Dr. Esseni is a member of the Technology Computer Aided Design Committee of the IEEE Electron Devices Society. He was member of the Technical Committee of the IEEE International Electron Devices Meeting, from 2003 to 2004. He is currently in the Technical Committee of the IEEE European Solid State Device Research Conference and the IEEE International Reliability Physics Symposium.



Michiel J. van Duuren was born in Groenlo, The Netherlands, in 1970. He received the M.S. degree in physics and the Ph.D. degree from the University of Twente, Enschede, The Netherlands, in 1993 and 1998, respectively.

After he graduated, he was with Philips Research Laboratories, Eindhoven, The Netherlands, where he started to work on reliability aspects, characterization methods, and physical modeling of floating gate and—since some years—SONOS memories. Since 2001, he has been with Philips Research Leuven (now NXP Semiconductors), Leuven, Belgium, where he is currently leading the “embedded memories” project.



Dusan S. Golubović received the M.S.E.E. degree from the University of Belgrade, Belgrade, Serbia, in 2000 and the Ph.D. degree in applied physics from Katholieke Universiteit Leuven, Leuven, Belgium, in 2005.

Since 2005, he has been a Senior Scientist with Philips Research Leuven (now NXP semiconductors), Leuven. His research interests are in the field of process integration and characterization of advanced embedded nonvolatile memories.



Nader Akil received the M.S. degree in physics from University of Reims Champagne-Ardenne, Reims, France, in 1995 and the Ph.D. degree from Metz University, Metz, France, in 1998. He spent the last year of his Ph.D. in the Department of Electrical and Computer Engineering, Vanderbilt University, Nashville, TN, working on light emission from silicon-based devices.

From 1999 to 2000, he was a Postdoctoral Associate with University of Twente, Enschede, The Netherlands, working on nanoscale silicon-based devices for applications in micrototal analysis systems. In 2000, he was with Philips Research Laboratories, Eindhoven, The Netherlands, and in 2001, he moved to Philips Research Leuven (now NXP Semiconductors), Leuven, Belgium, where he is currently working on modeling and characterization of floating gate and charge trapping memories.



Rob van Schaijk was born in Schaijk, The Netherlands, in 1971. He received the M.S. degree in physics from the Technical University of Eindhoven, Eindhoven, The Netherlands, in 1995, working on optical and transport properties of 2-D hole-gases in GaAs, and the Ph.D. degree from the University of Amsterdam, Amsterdam, The Netherlands, in 1999. His Ph.D. work was on magneto transport in low-dimensional semiconductor systems.

In 1999, he was with Philips Research, Eindhoven, and in 2000, he moved to Leuven, Belgium, to join Philips Research Leuven (now NXP semiconductors), which is in partnership with the Interuniversity Microelectronics Center. His research activities are in the field of process integration of embedded nonvolatile memories. In particular, the work is focused on integration of floating gate Flash and SONOS memory cells.

EXHIBIT D

Bit Cost Scalable Technology with Punch and Plug Process for Ultra High Density Flash Memory

H.Tanaka, M.Kido, K.Yahashi*, M.Oomura*, R.Katsumata, M.Kito, Y.Fukuzumi, M.Sato, Y.Nagata**,
Y.Matsuoka, Y.Iwata, H.Aochi and A.Nitayama

Center for Semiconductor Research & Development and Process & Manufacturing Engineering Center*,

TOSHIBA Corporation, Semiconductor Company

Toshiba Information Systems (Japan) Corporation**

8, Shinsugita-cho Isogo-ku, Yokohama Kanagawa 235-8522 Japan

Tel) +81-45-770-3274, Fax) +81-45-770-3210, E-mail) hiroyasu5.tanaka@toshiba.co.jp

Abstract

We propose Bit-Cost Scalable (BiCS) technology which realizes a multi-stacked memory array with a few constant critical lithography steps regardless of number of stacked layer to keep a continuous reduction of bit cost. In this technology, whole stack of electrode plate is punched through and plugged by another electrode material. SONOS type flash technology is successfully applied to achieve BiCS flash memory. Its cell array concept, fabrication process and characteristics of key features are presented.

Introduction

It is strongly demanded to keep a trend of increasing bit density and reducing bit cost of flash memories. To achieve that, aggressive scaling of the device dimension and multi-level-cell have been developed. However, bit cost will turn to rise up near future, as shown in figure 1, because process cost will increase more rapidly than shrink rate.

One solution to avoid such situation is three dimensionally stacked array structures as reported in previous papers [1] [2], in which planar NAND structures are simply stacked. But there are two severe issues on those structures. One is that several critical lithography steps to make minimum feature size are necessary for each stacked layer. The other is that the area of driver transistors for control gates is multiplied by the number of layers. Dotted line in the figure 2 illustrates the estimated bit cost of three dimensionally stacked NAND, calculated by a simple formula inset in figure 2. It is important to note that the cost does not decrease or even increases due to a yield loss and an area penalty if the number of stacked layers is more than three.

In this paper, a novel SONOS flash memory using a newly devised Bit-Cost Scalable (BiCS) technology which realizes a multi-stacked array with a few constant number of critical lithography process steps, without area penalty and with continuous reduction of bit cost as shown as solid line in figure 2 is proposed.

Concept

Figures 3(a), 3(b) and figure 4 show schematic Birds-eye view, top down view of BiCS flash memory array and the equivalent circuit, respectively. String of the cells is on the plugs located vertically in the holes punched through whole stack of the gate plates. Each plate acts as control gate except the lowest plate which takes a role of the lower select gate. Cell size of BiCS memory is $6F^2/n$, where n is the number of control gate plates. Bit density can be increased by adding more gate plates, while the number of the critical lithography steps remains constant because whole stack of control gates is punched with only one lithography step. A single cell is accessed by control gate on the string which is selected by a bit line and an row select line. The bottom of memory plug is connected to source diffusion formed on the silicon substrate. For the erase operation of BiCS Flash memory, hole current which is generated by GIDL near the lower select gate is used. Edges of control gate are processed to be stair-like

structure for via holes which are connected to driver transistors. They are shared by several rows of strings, so that the area for driver transistors does not increase even if more control gates are stacked. Table 1 summarizes the comparison of BiCS Flash and simply stacked 3D NAND.

Fabrication

Figure 5 shows the fabrication sequence. Lower select gate transistor, memory string and upper select gate transistor are fabricated individually. Gate material is P+ poly-Si. Holes for transistor channel or memory plug are punched through and LPCVD TEOS film or ONO films are deposited. The bottom of dielectric films are removed by RIE and plugged by amorphous Si. Arsenic is implanted and activated for drain and also source of upper device. Figure 6 shows schematic cross sections of vertical FET and SONOS memory cell on poly-Si gate. It should be pointed out that ONO films are deposited in the opposite order of the conventional SONOS device, i.e. from LPCVD TEOS film as top block oxide (5nm), LPCVD SiN film (11nm) and LPCVD TEOS film as tunnel oxide (2.5nm). Edges of control gate are processed into stair-like structure by repeating of RIE and resist sliming as shown in figure 8. For minimizing disturb, whole stack of control gate and lower select line are etched to have a slit which separates a block of memory plugs from each other. Only upper select gate is cut into line pattern to work as row address selector. Via hole and BL are processed on the array and peripheral circuit simultaneously.

Result and Discussion

Figures 7(a) and 7(b) show the cross sectional SEM images of the successfully fabricated 90nm BiCS flash cell array and the stair-like structure at the edge of control gates. Figure 7(c) is a top down photograph of $n \times 512k$ bit BiCS Flash macro.

Id-Vg characteristics of select gate FET ($L=150nm$, $\Phi=90nm$) is shown in figure 9. Subthreshold slope is 250 mV/dec. In the case that I_{off} is 30 pA, on-current is 2.4 μA . Figure 10 shows Id-Vg characteristics of the vertical SONOS FET after program and erase. Figure 11 shows the program/erase characteristics of SONOS FET. Endurance to $1E5$ cycles is shown in figure 12. Vth shift is less than 0.5V after $1E4$ cycle. Data retention measured on SONOS FET at initial and after 1000 cycles of endurance are shown in figure 13. 2.5V of Vth window is remains after 10years.

Conclusion

We have proposed BiCS flash memory as one of the most cost-effective memories to increase bit density. We have confirmed that this technology is a promising candidate for the Ultra High Density Memory.

Reference

- [1] Soon -Moon Jung et al, IEDM Tech. Dig., pp 37-40, 2006.
- [2] Erh-Kun Lai et al, IEDM Tech. Dig., pp 41-44, 2006.

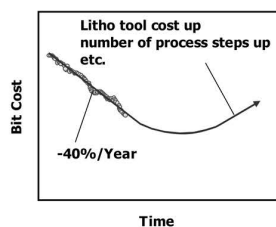
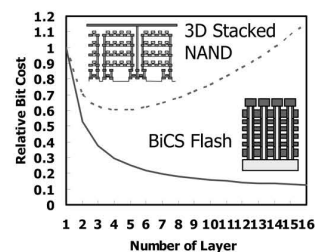


Fig. 1 Bit cost trend of flash memory.

TABLE I
Comparison of three dimensional flash memories.

	Pros	Cons
3D Stacked NAND	Expansion of Current Technology	Higher Cost (Large number of Lithography Steps)
BiCS Flash	Lower Cost (Small number of Lithography Steps)	New Technology - LPCVD tunnel oxide - Vertical FET - Multi Gate Stack RIE



$$\frac{1}{n} (C_f + nC_v) \left(\frac{1+A}{1-Y} \right)^{n-1}$$

n : Number of stacked layers.
 C_f : Cost for common part.
 C_v : Cost per single layer.
 A : Area penalty rate per single layer.
 Y : Yield loss per single layer.

Fig. 2 Bit Cost scalability of three dimensional flash memory.

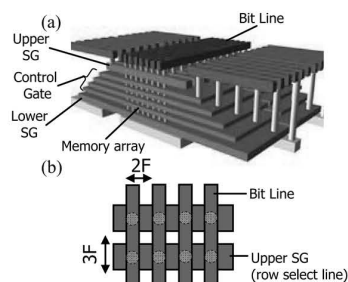


Fig. 3 (a) Birds-eye view of BiCS flash memory, (b) Top down view of BiCS flash memory array.

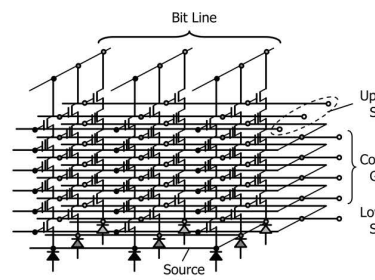


Fig. 4 Equivalent circuit of BiCS flash memory.

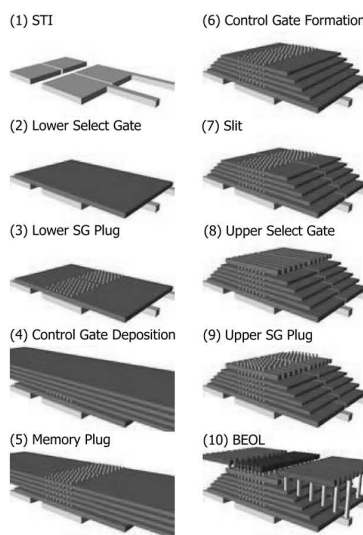


Fig. 5 Fabrication sequence of BiCS flash memory.

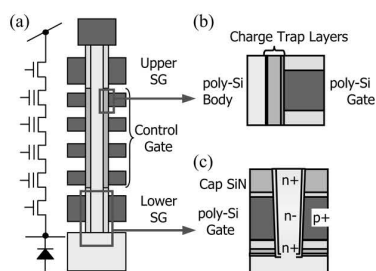


Fig. 6 (a) Cross section of BiCS flash memory string, (b) Cross section of vertical SONOS cell, (c) Cross sections of vertical FET.

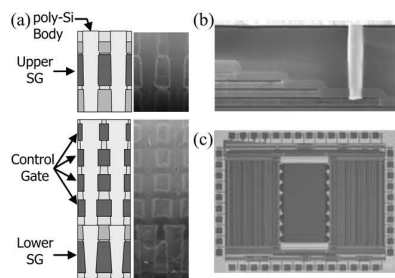


Fig. 7 (a) Cross sectional SEM of BiCS flash memory string, (b) Cross sectional SEM of edge of control gates, (c) n x 512 kbit macro image.

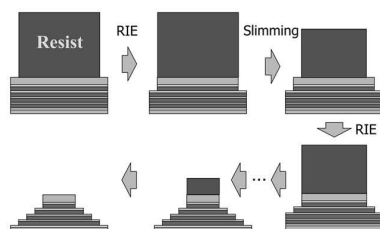


Fig. 8 Fabrication Sequence of edge of control gates into stair-like structure.

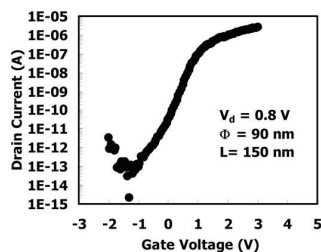


Fig. 9 Id-Vg characteristics of vertical transistor.

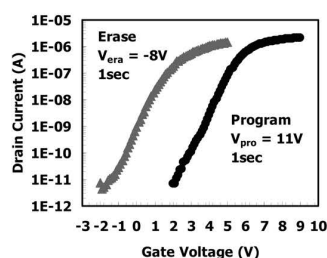


Fig. 10 Id-Vg characteristics for Program/Erase of vertical SONOS.

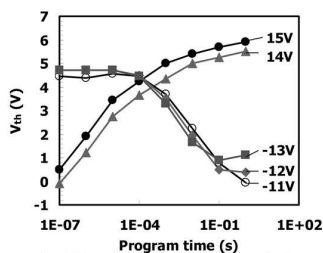


Fig. 11 Program/Erase characteristics.

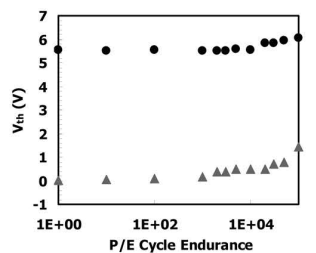


Fig. 12 Endurance of vertical SONOS.

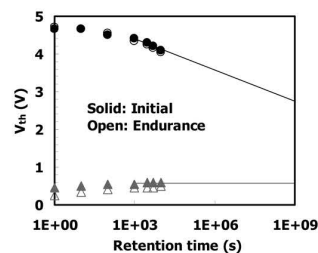


Fig. 13 Data retention of vertical SONOS.

Ex. A Abstract

VDNROM: A Novel Four-Bits-Per-Cell Vertical Channel Dual-Nitride-Trapping-Layer ROM for High Density Flash Memory Applications

Publisher: IEEE

Cite This

 PDF

Falong Zhou ; Yimao Cai ; Ru Huang ; Yan Li ; Xiaonan Shan ; Jia Liu

All Authors

1
Cites in
Paper

4
Cites in
Patents

155
Full
Text Views



Abstract	Abstract: A novel vertical channel nonvolatile memory cell with oxide-nitride-oxide-nitride-oxide (ONONO, dual nitride trapping layers) dielectrics stack is proposed and experimentally demonstrated for the first time. Compared with the conventional planar NROM cell, since the cell area of the proposed vertical structure is independent of the gate length, the VDNROM structure can relax the limitation of the gate length scaling, and can have high capability of cell area shrinking. The fabrication process of this VDNROM device is basically compatible with planar CMOS technology. The VDNROM cell can be programmed and erased by the hot carrier injection to the localized trapping dual-nitride-layers, so it can achieve a four physical bits storage capability each cell. The reliability behaviors including the cycling endurance and the bake retention at 150degC have also been investigated and show the acceptable characteristics. The experiment results verify the VDNROM cell as a good candidate for high-density applications	
Document Sections I. Introduction II. Device Structure and Fabrication III. Operation Mechanism and Experiment Results IV. Conclusion	Published in: 2006 European Solid-State Device Research Conference Date of Conference: 19-21 September 2006 Date Added to IEEE Xplore: 12 February 2007 Print ISBN:1-4244-0301-4 DOI: 10.1109/ESSDER.2006.307679 Publisher: IEEE Conference Location: Montreux, Switzerland	
Authors Figures References  Citations		

[Keywords](#)[Metrics](#)[More Like This](#)[▼ ISSN Information:](#)

SECTION I.

Introduction

Recently, SONOS flash memory has been paid more attention and considered as a promising technology bridging the gap between the floating-gate flash and the next generation memories because of its intrinsic better endurance, good scalability, simpler fabrication process and its perfect compatibility with conventional CMOS technology for embedded applications [1]– [3]. However, conventional SONOS cell is only one-bit-per-cell storage, and the data retention is still a big problem due to the direct tunneling of stored charge through the ultra thin tunneling oxide.

The NROM concept [4]– [5] was proposed to solve the drawbacks of SONOS by using channel hot electrons injection (CHEI) for programming and band-to-band tunneling induced hot holes injection (BBTIHH) for erasing. This device can allow thicker tunneling oxide, while still maintaining high efficient Program/Erase (P/E) operations. Moreover, the P/E mechanisms lead to localized charge storage in the trapping nitride layer, so the NROM can store two distinct bits per cell. However, the further scaling of the cell area of the conventional planar NROM device for higher density is limited by the gate length scaling for two reasons. Firstly, the drain voltage can not be scaled below 3.2V for CHEI, and therefore the punch-through effect would become worse as the gate length scaling [6]. Secondly, since two bits are stored in physically distinct areas in the nitride layer, the lateral distributions of the charges have to be well separated to avoid any interference between the two bits [7].

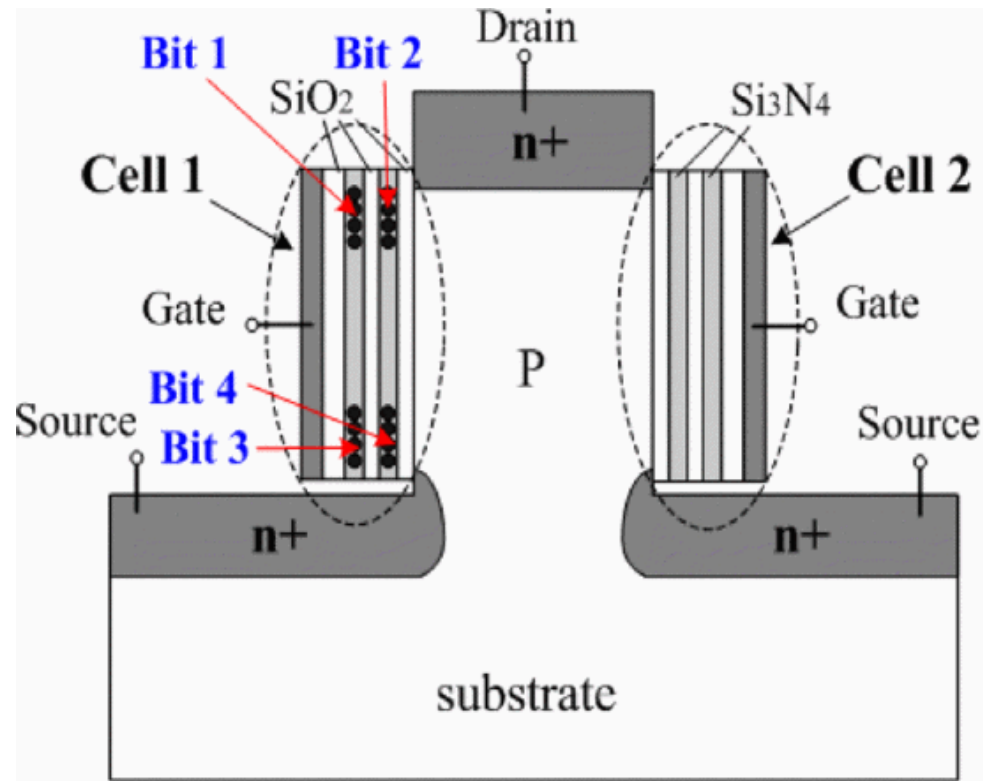


Fig. 1 Schematic cross sectional view of the proposed 4 bits/cell VDNROM flash memory.

In this work, as shown in [Fig. 1](#), we proposed and demonstrated experimentally a novel vertical channel dual-nitride-trapping-layers ROM (VDNROM) structure nonvolatile memory device achieving four-bits-per-cell function for the first time. Compared with the conventional planar NROM cell, since the cell area of this vertical channel structure is independent of the gate length, the VDNROM structure can relax the limitation of the gate length scaling, and can have high capability to shrink the cell area. Moreover, the fabrication of this VDNROM device is basically compatible with planar CMOS technology. The experimental results demonstrate continued better scalability of the VDNROM cell for high-density applications.

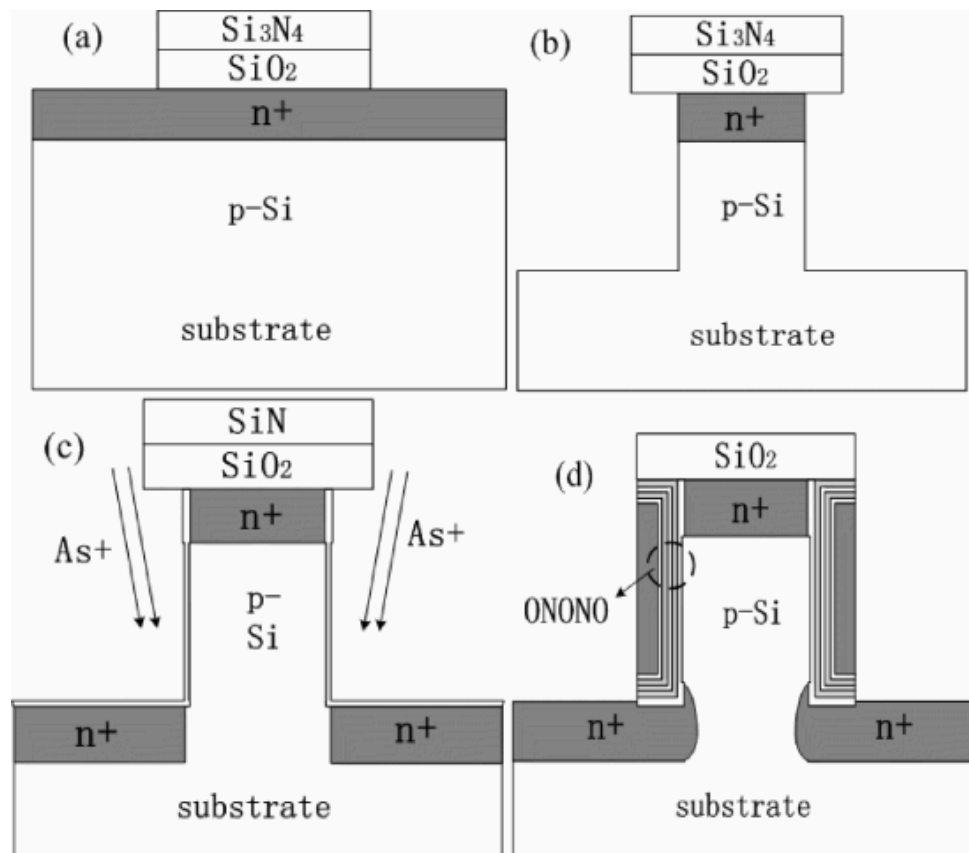


Fig. 2. The process flow of the proposed 4bits/cell VDNROM flash memory.

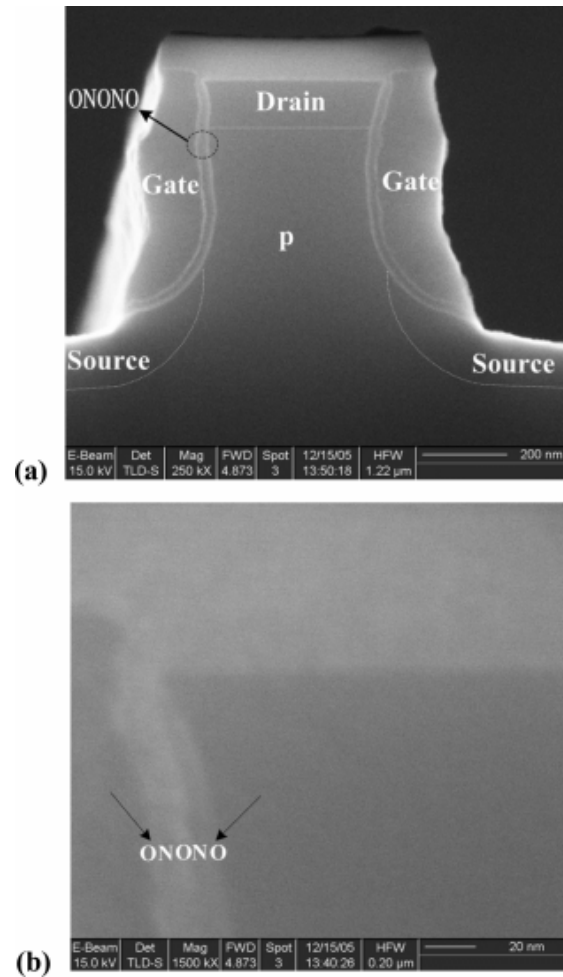


Fig. 3. The SEM images of the proposed 4bits/cell VDNROM flash memory cell (a) and the ONONO gate dielectric stack (b).

SECTION II.

Device Structure and Fabrication

The schematic cross-sectional view of the proposed VDNROM flash memory cell with the oxide-nitride-oxide-nitride-oxide (ONONO) gate dielectric stack is illustrated in [Fig. 1](#). The charges can be stored in the dual nitride trapping layers above the channel next to the n+/p junctions, so this VDNROM cell can achieve a four physical bits storage capability as shown in [Fig. 1](#). The tunneling oxide and the block oxide are thicker than 5 nm to avoid any direct tunneling of the charges, and the gate dielectric stack of the ONONO five layers (the tunneling oxide, the first trapping nitride layer, the inter-nitride oxide, the second trapping nitride layer, the block oxide) are 7/4/5/5/13 nm respectively.

Referring to [Fig. 1](#), the process flow to realize such a 4bits/cell VDNROM device was shown in [Fig. 2\(a\)](#)-(b). After isolation, the channel boron doping and the drain arsenic doping were implemented. Two layers of oxide and nitride was deposited, patterned and etched to form the dual-layers hard masks as shown in [Fig. 2\(a\)](#). The silicon substrate was then etched by ICP (Inductive Coupled Plasma) to form the steep device mesa. Next, the device mesa was etched isotropically by RIE (Reactive Ion Etch) to form the recessed self-aligned poly gate structure shown in [Fig. 2\(b\)](#). To remove the surface damage on the sidewalls of the device mesa, thin sacrificial oxide was grown, which was followed by source arsenic implantation ($4 \times 10^{15} \text{cm}^{-2}$, 100KeV), as shown in [Fig. 2\(c\)](#). After sacrificial oxide removed, 7.0-nm gate oxide was grown at 850°C by dry thermal oxidation. Next, a dielectric stack of trapping-nitride/inter-oxide/trapping-nitride/block-oxide (NONO) with 4/5/5/13 nm respectively was deposited. Then, a polysilicon film was deposited, implanted and etched to form the self-aligned gate spacer. After the nitride hard mask removed, rapid thermal annealing was implemented to activate the dopants in the S/D region and poly gate, as shown in [Fig. 2\(d\)](#). Finally, the device was covered by low temperature deposited oxide (LTO), and followed by the rest of the process such as contact opening, metal defining and annealing. [Fig. 3\(a\)](#) shows the SEM (scanning electron microscope) micrograph of the cross section of the fabricated VDNROM cell, and the SEM image of the ONONO gate dielectric stack is shown in [Fig. 3\(b\)](#).

SECTION III.

Operation Mechanism and Experiment Results

The VDNROM cell is programmed and erased by channel hot carriers injection. The four states (“00”, “01”, “10”, “11”) corresponding to bit-1 and bit-2 are shown in [Fig. 4\(a\)](#)-(d), and the programming and erasing operations of the four states are shown in [Table I](#). For example, applying 7.0 V on the drain electrode, 12.0 V on WL (the gate electrode) and grounding the source electrode for programming the state “00” in [Fig. 4\(a\)](#). The accelerated channel electrons are injected into the both two nitride layers of the ONONO gate dielectric stack near the drain

junction, a self-aligned storage with the junction edge. As shown in [Fig. 4\(b\)](#) and [Table I](#), the state of “01” is programmed by the hot electrons injection to the first trapping nitride layer with lower drain voltage 5.0 V (compared with the programming of the state of “00”), 6.0 V on WL (the gate electrode) and grounding the source electrode. The state of “10” shown in [Fig. 4\(c\)](#) is realized by two steps. As shown in [Table I](#), the first step is the hot electrons injection to the dual trapping nitride layers with $V_g=12.0$ V, $V_d=7.0$ V and $V_s=0$ V, and the second is the hot holes injection to the first nitride layer applying $V_d=-8.0$ V, $V_g=5.0$ V and $V_s=0$ V. Holes are generated by band-to-band tunneling in the deep depletion layer under the gate dielectric stack above the n^+/p junction [\[4\]](#), [\[8\]](#). Holes increase their energy due to the lateral field in depletion region of the channel and they are injected through the bottom oxide into the nitride trapping layer, where they recombine with the stored electrons or are trapped. As shown in [Fig. 4\(d\)](#), the state of “11” is the initial state and it can also be realized by erasing from other states shown in [Table I](#).

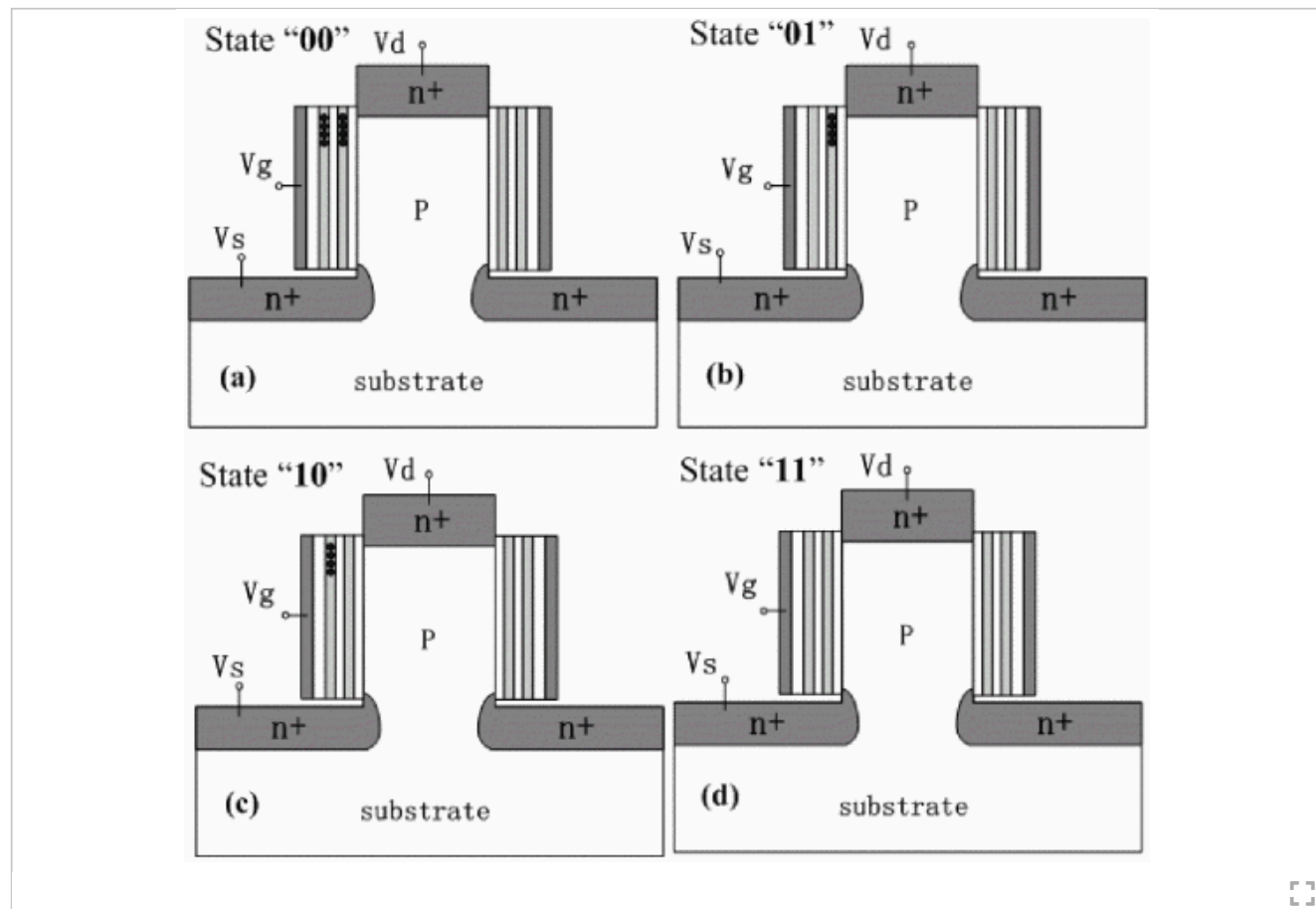
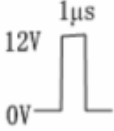
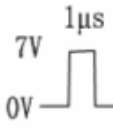
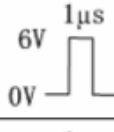
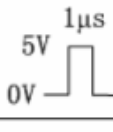
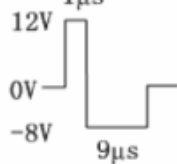
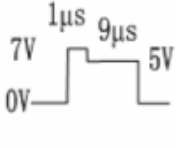
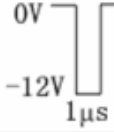
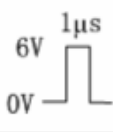


Fig. 4. The four states corresponding to bit-1 and bit-2 of the proposed VDNROM flash memory.

TABLE I. VDNROM Flash Memory Program/Erase Operating Conditions (Four states corresponding to bit-1 and bit-2)

State	P/E	Gate (Pulse)	Drain (Pulse)	Source	Time	Vth
'00'	Program from "11"			0 V	10 ms	5.5V
'01'	Program from "11"			0 V	10 ms	4.3V
'10'	Program from "11"			0 V	100 ms	3.3V
'11'	Erase			0 V	100 ms	2.3V

[Fig.5](#) shows transfer characteristics of the VDNROM flash memory cell with 0.3 μ m channel length in the four states. The threshold voltage of State "11" is 2.3V (the same Vth as the initial). State "00" has a threshold shift of $\Delta V_{th} = 3.2V$ from State "11". The ΔV_{th} of 2.0V in State "01" is observed. The threshold voltage of State "10" is 3.3V, and the ΔV_{th} is 1.0V. The transfer characteristics of each state exhibits excellent subthreshold performance as shown in [Fig.5](#). The programming and erasing characteristics of the VDNROM flash memory are shown in [Fig.6](#).

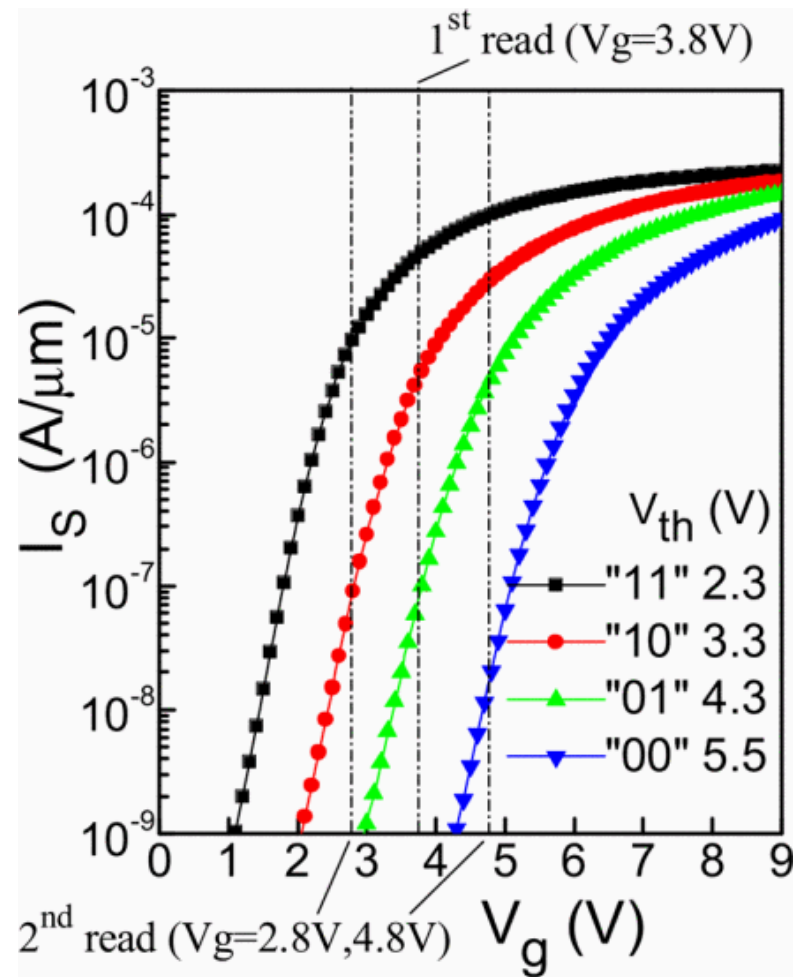


Fig. 5. The transfer characteristics of the programmed/erased bit-1 and bit-2 of the VDNROM flash memory.

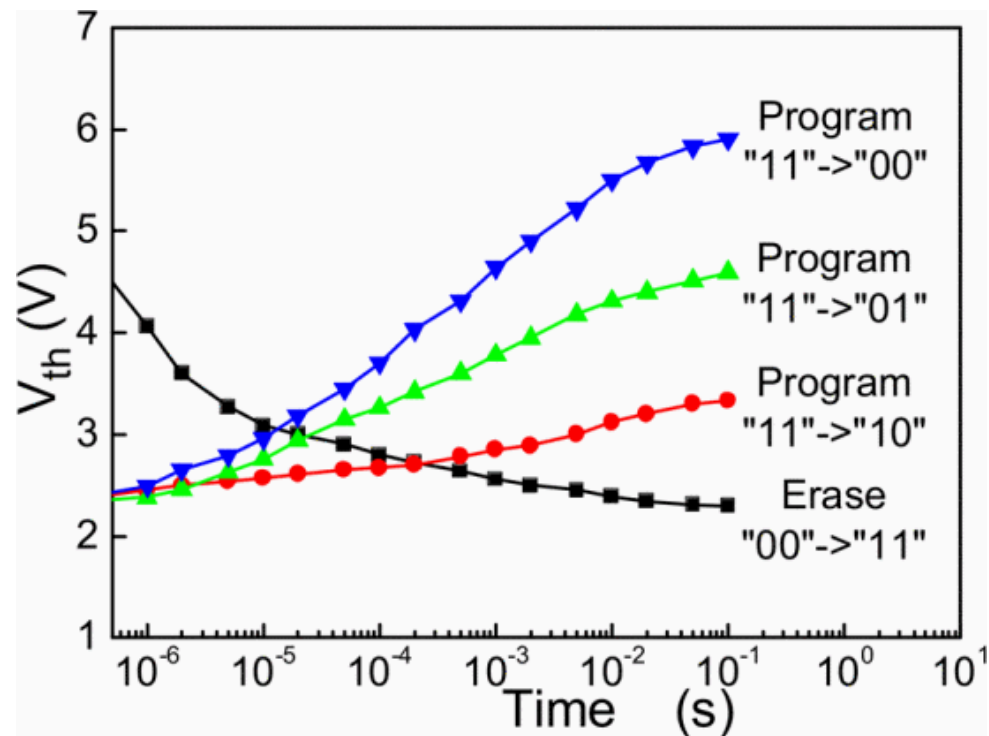


Fig. 6. The Programming/Erasing characteristics of the VDNROM flash memory.

The device is read in the “reverse” direction compared to programming, by applying the read biases to source electrode (1.5V) and grounding the drain electrode. This reverse read mode can enhance the threshold voltage sensitivity to the localized trapping charges [4]. As shown in Fig. 5, during the first read operation, the “00” and “01” states can be distinguished from the “10” and “11” states at 3.8V (gate voltage). After that, State “11” from State “10” and State “01” from State “00” at 2.8V and 4.8V (gate voltage), respectively, is performed for the second read operation.

Similarly, the bit-3 and bit-4 can be programmed, erased and read by interchanging the role of the drain/source electrodes.

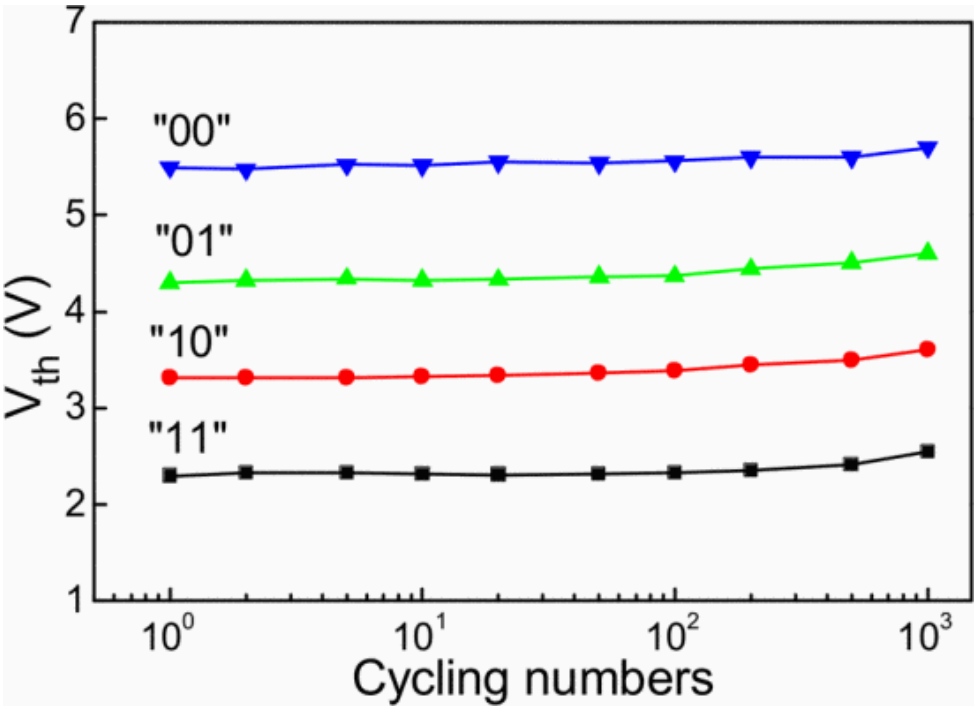


Fig. 7. The Endurance characteristics of the VDNROM flash memory.

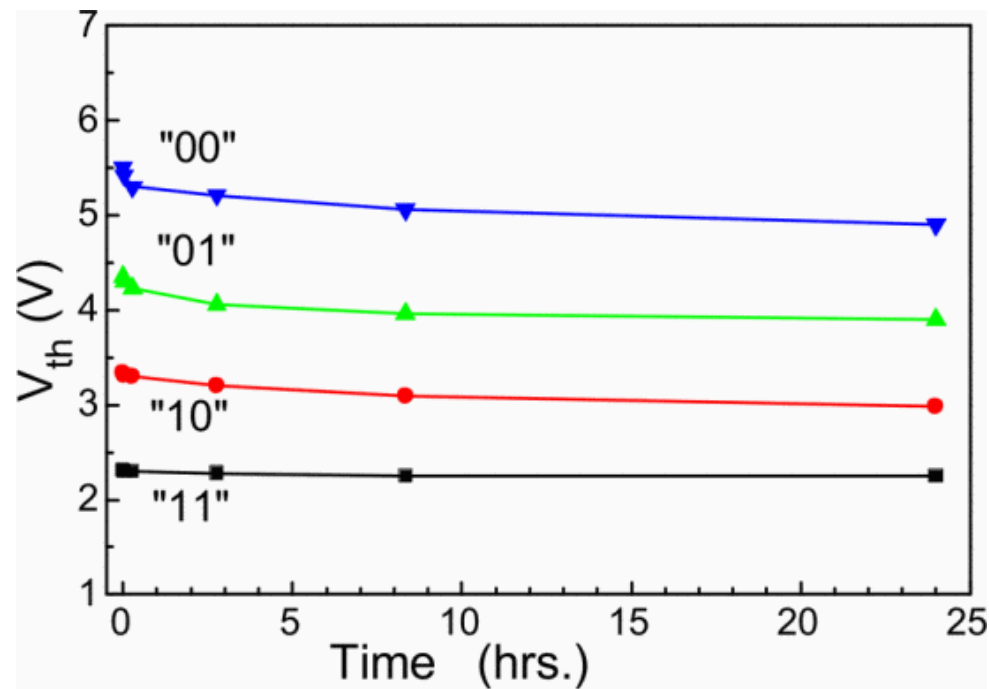


Fig. 8. The bake retention performance of the VDNROM flash memory at 150°C

[Fig. 7](#) shows the endurance characteristics of the four states (corresponding to bit-1 and bit-2). The 2-bit cycling loop consists of the sequence of bit-2 program (State “11” to State “01”) → bit-1 program (“01” to “00”) → bit-2 erase (“00” to “10”) → bit-1 erase (“10” to “11”) that can be considered as the worst case, because the tunneling oxide has been tunneled by twice hot electrons and twice hot holes in one cycle (the inter-nitride oxide by once electrons and once holes). After 10^3 P/E cycles, the worst V_{th} window of 0.8V is maintained in spite of the V_{th} increase at the erased cell. These electrical results indicate that the cycling endurance performance of the VDNROM cell is acceptable.

[Fig. 8](#) shows the bake retention performance of the VDNROM measured at 150°C. 0.4V small degradation and saturation of V_{th} are observed after 24 hours baking measurement. With the better quality tunneling oxide, the smaller degradation of V_{th} window can be observed.

SECTION IV.

Conclusion

For the first time, a novel VDNROM flash memory cell with an oxide-nitride-oxide-nitride-oxide (ONONO) gate dielectric stack has been proposed, demonstrated experimentally, and successfully operated in four-bits-per-cell mode in the 0.3 μ m channel length vertical device. In the conventional planar NROM cell, the cell area further scaling is limited by the gate length scaling. In contrast, since the cell area of the VDNROM structure is independent of the gate length, the structure can have high capability of the cell area shrinking with the relaxed requirement of the gate length scaling. Moreover, the fabrication of the VDNROM is basically compatible with planar CMOS technology. The reliability performances including the cycling endurance and the bake retention at 150°C have also been investigated and are acceptable. This new technology can be used for data storage due to the high density, and for embedded applications due to the simplicity of the process.

ACKNOWLEDGMENT

The authors would like to thank the Microfabrication Laboratory staffs of Peking University for their supports in device fabrication. This work is supported by National Natural Science Foundation of China (No. 90207004) and State Key Fundamental Research Project of China.

Authors	▼
Figures	▼
References	▼
Citations	▼
Keywords	▼
Metrics	▼

[Back to Results](#)

IEEE Personal Account

CHANGE
USERNAME/PASSWORD

Purchase Details

PAYMENT OPTIONS
VIEW PURCHASED
DOCUMENTS

Profile Information

COMMUNICATIONS
PREFERENCES
PROFESSION AND
EDUCATION
TECHNICAL INTERESTS

Need Help?

US & CANADA: +1 800
678 4333
WORLDWIDE: +1 732
981 0060
CONTACT & SUPPORT

Follow



[About IEEE Xplore](#) | [Contact Us](#) | [Help](#) | [Accessibility](#) | [Terms of Use](#) | [Nondiscrimination Policy](#) | [IEEE Ethics Reporting](#)  | [Sitemap](#) | [IEEE Privacy Policy](#)

A public charity, IEEE is the world's largest technical professional organization dedicated to advancing technology for the benefit of humanity.

© Copyright 2025 IEEE - All rights reserved, including rights for text and data mining and training of artificial intelligence and similar technologies.

Ex. B Abstract

A Multi-Layer Stackable Thin-Film Transistor (TFT) NAND-Type Flash Memory

Publisher: IEEE

Cite This

PDF

Erh-Kun Lai ; Hang-Ting Lue ; Yi-Hsuan Hsiao ; Jung-Yu Hsieh ; Chi-Pin Lu ; Szu-Yu Wang

All Authors

59
Cites in
Papers

47
Cites in
Patents

2455
Full
Text Views



Abstract

Document Sections

I. Introduction

II. Fabrication

III. Device Characterization

IV. Comparison between the Top and Bottom Layers

V. NAND Array Functions

Show Full Outline ▼

Authors

Figures

References

Citations

Abstract:

A double-layer TFT NAND-type flash memory is demonstrated, ushering into the era of three-dimensional (3D) flash memory. A TFT device using bandgap engineered SONOS (BE-SONOS) (Lue et al., 2005, Lai et al., 2006) with fully-depleted (FD) poly silicon (60 nm) channel and tri-gate P⁺-poly gate is integrated into a NAND array. Small devices (L/W=0.2/0.09 μm) with excellent performance and reliability properties are achieved. The bottom layer shows no sign of reliability degradation compared to the top layer, indicating the potential for further multi-layer stacking. The present work illustrates the feasibility of 3D flash memory

Published in: 2006 International Electron Devices Meeting

Date of Conference: 11-13 December 2006

DOI: 10.1109/IEDM.2006.346903

Date Added to IEEE Xplore: 16 April 2007

Publisher: IEEE

► ISBN Information:

Conference Location: San Francisco, CA, USA

▼ ISSN Information:

SECTION I.

[Keywords](#)[Metrics](#)[More Like This](#)

Introduction

Recently, the NAND flash becomes very popular for portable electronics, and the demand for memory density multiplies every year. Although NAND Flash is aggressively scaled [3, 4], it becomes more and more difficult to follow Moore's law to continue the scaling due to both process and device limitations. Three-dimensional (3D) multi-layer-stack memory is another path to ultra-high density memory.

We demonstrated a low-thermal budget TFT NAND device recently [2]. The TFT memory employs BE-SONOS [1] device, which provides both efficient hole tunneling erase and good data retention [1]. Moreover, BE-SONOS uses only matured front-end materials (oxide, nitride and poly gate). This greatly enables the multi-layer stack capability without metal contamination issues. In this work, we further fabricate a double-layer TFT NAND Flash. The performance and reliability of bottom and top layers are examined to provide guidelines for further multi-layer stacking. The effects of device dimension, reliability, and NAND array operation are addressed.

SECTION II. Fabrication

The typical device parameters and fabrication flow are shown in Table 1 and Table 2. Figures 1– 4 show the cross-sectional views of the double layer TFT NAND devices. The TFT device has a 60 nm thick poly silicon channel over the buried oxide. The poly silicon is an amorphous silicon (a-Si) layer deposited by low-pressure chemical vapor deposition (LPCVD) process, followed by a low-temperature thermal annealing (600 °C) for crystallization. A multi-layer O₁/N₁/O₂ tunneling dielectric is used to achieve easy hole tunneling during erase, while eliminating the direct tunneling leakage during retention [1]. Next, SiN trapping layer (N₂) and top blocking oxide (O₃) are grown. A heavily doped p⁺-poly gate is adopted to suppress the gate injection during – FN erase. The multi-layer ONONO are shown in Fig. 1. The device has a tri-gate structure as shown in Fig. 2.

The top oxide process has the largest thermal budget. Two top oxide (O₃) formation processes are compared: an LPCVD oxide (HTO) with rapid thermal annealing, and an in-situ steam generation (ISSG) oxidation to convert a portion of the trapping nitride (N₂) into oxide.

After gate patterning, shallow junction is formed by low-energy As⁺ implantation. Planarization is then carried out by HDP oxide deposition and CMP. After forming the bottom-layer TFT devices, the processes are duplicated to form the top-layer TFT devices. Contact etching for the two layers are performed independently to avoid over etching. The final double-layer TFT NAND Flash memory is shown in [Figs.3– 4](#).

SECTION III.

Device Characterization

(a) Short-Channel Effect

The I - V and V_T for the memory cells with various channel lengths are shown in [Figs.5, 6](#), respectively. Similar to bulk devices, the transconductance (g_m) increases with shorter channel lengths. [Figure 6](#) shows that the-FN erased state (-17 V, 50 msec) follows the same V_t -roll-off characteristics with the initial state, but with 3 V lower V_t . ISSG and HTO top oxide devices show similar V_t roll-off trends.

(b) Narrow-Width Effect

The I - V characteristics for memory cells with various channel widths are shown in [Fig.7](#). Generally a narrower device shows a better subthreshold slope (S.S.). [Figure 8](#) shows that for a narrow device the transconductance per width (g_m/W) increases, while V_t decreases. Thus a narrower device has a better DC performance due to better gate control capability of the tri-gate structure.

(c) Program/Erase Characteristics

The device is programmed by +FN electron injection ([Fig.9](#)), and erased by –FN hole tunneling ([Fig.10](#)). p⁺-poly gate is used to suppress the gate electron injection during – FN erase. Accordingly, the device can be erased into depletion mode. Both +FN program and –FN erase saturate for each bias voltage. Samples with ISSG and HTO top oxide show similar P/E speed.

(d) P/E Cycling Endurance

The P/E cycling endurance is shown in [Fig.11](#). Erase degradation occurs at < 1K dumb-mode cycles. ISSG top oxide devices show better endurance than those with HTO top oxide. The top oxide process has large impact on

the cell endurance because top oxide suffers a large electric field stress during P/E, while the top oxide has lower quality due to the lower thermal process.

The endurance of TFT device is understandably inferior to the bulk BE-SONOS device [1, 5], where a high-temperature (950~1000 °C) process is used to convert nitride into top oxide. The reason may be attributed to that the lower temperature ISSG and HTO top oxides have lower quality that easily trap charges after P/E cycle stressing, resulting in lower endurance.

Adjusted P/E cycling (as commonly practiced) achieves better endurance (Fig. 12). The corresponding P/E shot numbers are recorded in Fig. 13. After 1K cycling the device starts to degrade and more P/E shots are needed to maintain the V_t levels. Figure 14 illustrates that ISSG top oxide has a better subthreshold slope than HTO top oxide. Typical IV characteristics during 10K P/E cycling are shown in Fig. 15.

SECTION IV.

Comparison between the Top and Bottom Layers

Figure 16 compares the V_T roll-off of erase saturation state for the bottom and top layers. The bottom layer has a slightly faster V_T -roll-off. In addition, the V_T distribution of the bottom-layer is wider. This is because the bottom layer undergoes a higher thermal budget. For better bottom layer performance thermal budget should be further reduced.

The P/E cycle endurance, S.S. degradation, and retention are shown in Figs. 17, 18 and 19, respectively, for the short-channel device ($L_g/W=0.2/0.09$ μm). The bottom layer shows no sign of reliability degradation compared to the top layer. This indicates that further multi-layer stack is feasible.

SECTION V.

NAND Array Functions

The reading method for the NAND array is shown in [Fig. 20](#). Since TFT NAND array has higher series resistance compared to the bulk device, read current saturates at large gate overdrive. The read current is enhanced by increasing V_{PASS} . For $V_{\text{PASS}} = 7 \text{ V}$, and $V_{\text{READ}} = 1 \text{ V}$, the read current for a 16-string NAND array approaches $1 \mu\text{A}$ ample for data Flash design.

The program-inhibit characteristics of TFT NAND array are illustrated in [Fig. 21](#). Cells C and D are under medium-field gate disturb, while cell B can be program-inhibited by raising the channel potential.

SECTION VI.

Summary

For the first time, a successful double-layer TFT NAND Flash memory is demonstrated. The experimental results are very encouraging for future 3D Flash memories.

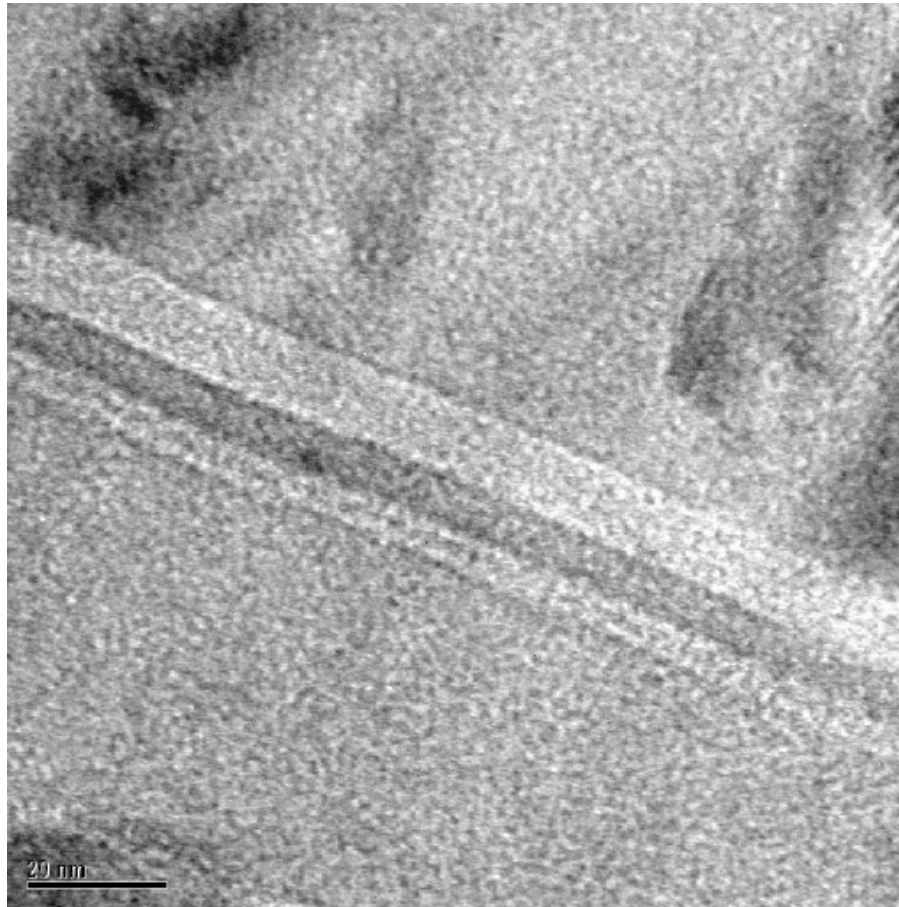


Fig. 1 TEM of the O1/N1/O2/N2/O3 structure for TFT BE-SONOS.

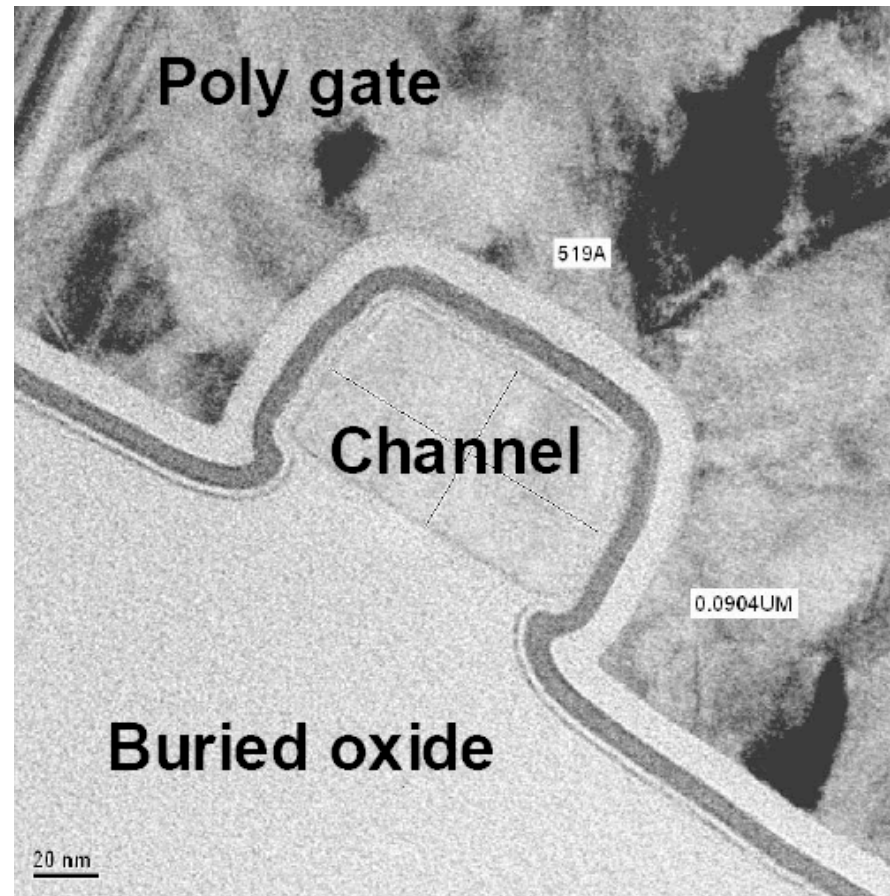


Fig. 2 Tri-gate structure of the TFT device.

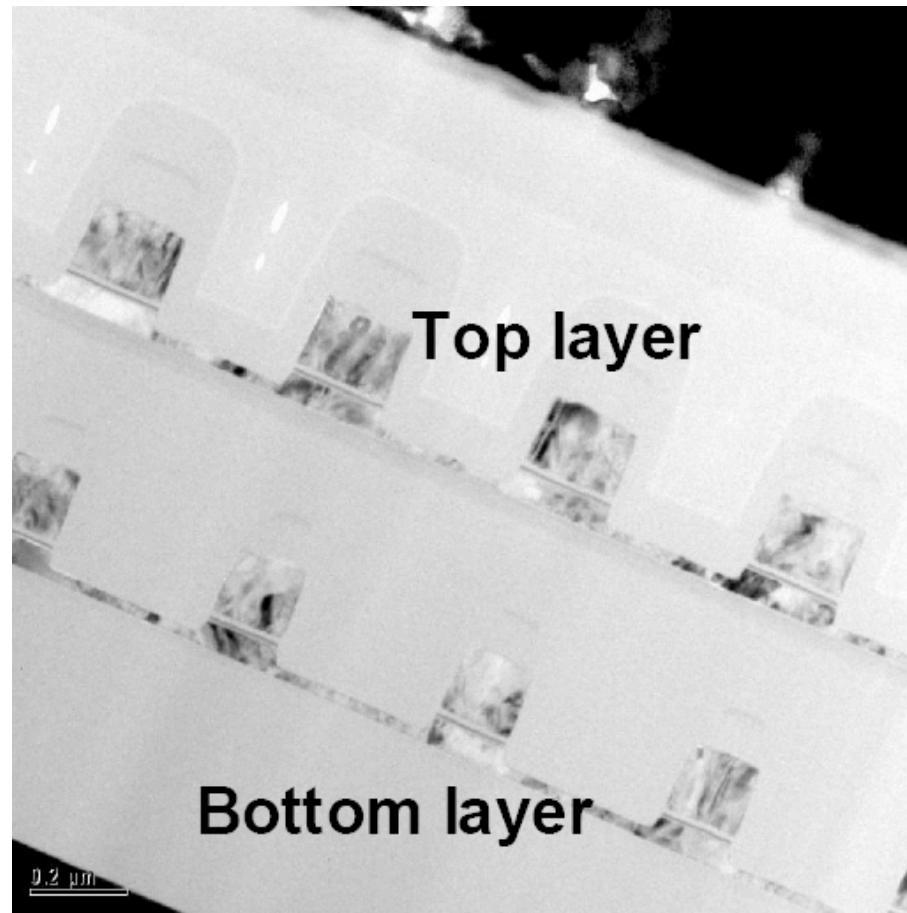


Fig. 3 Channel-length direction of double-layer TFT NAND devices.

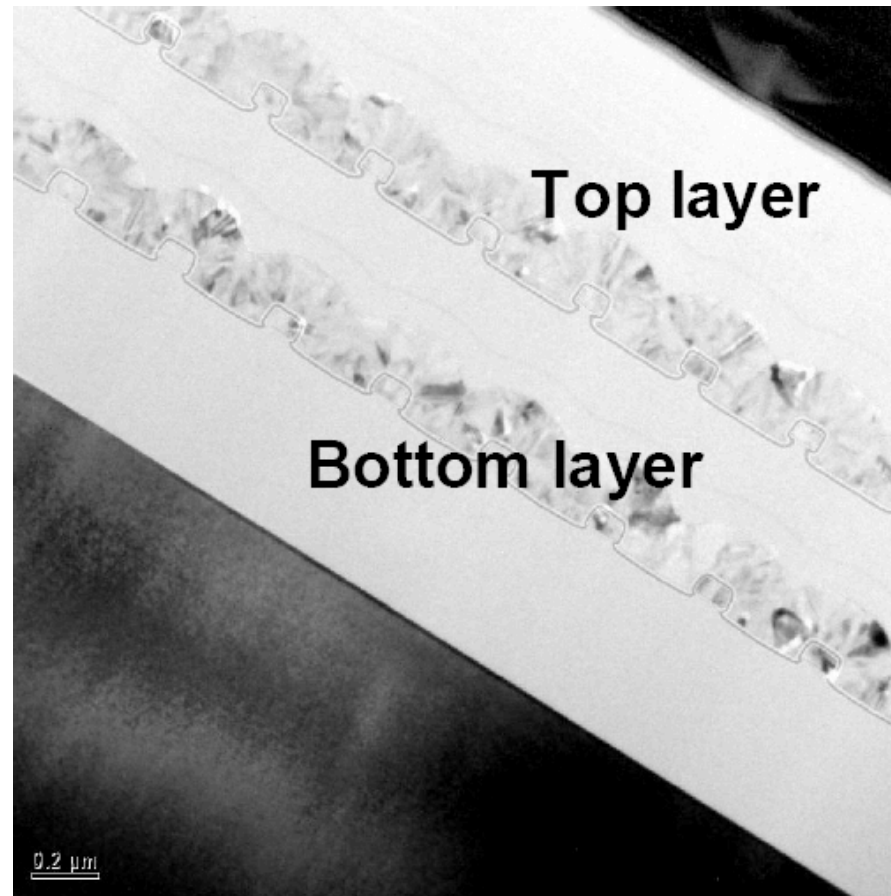


Fig. 4 Channel-width direction of double-layer TFT NAND devices.

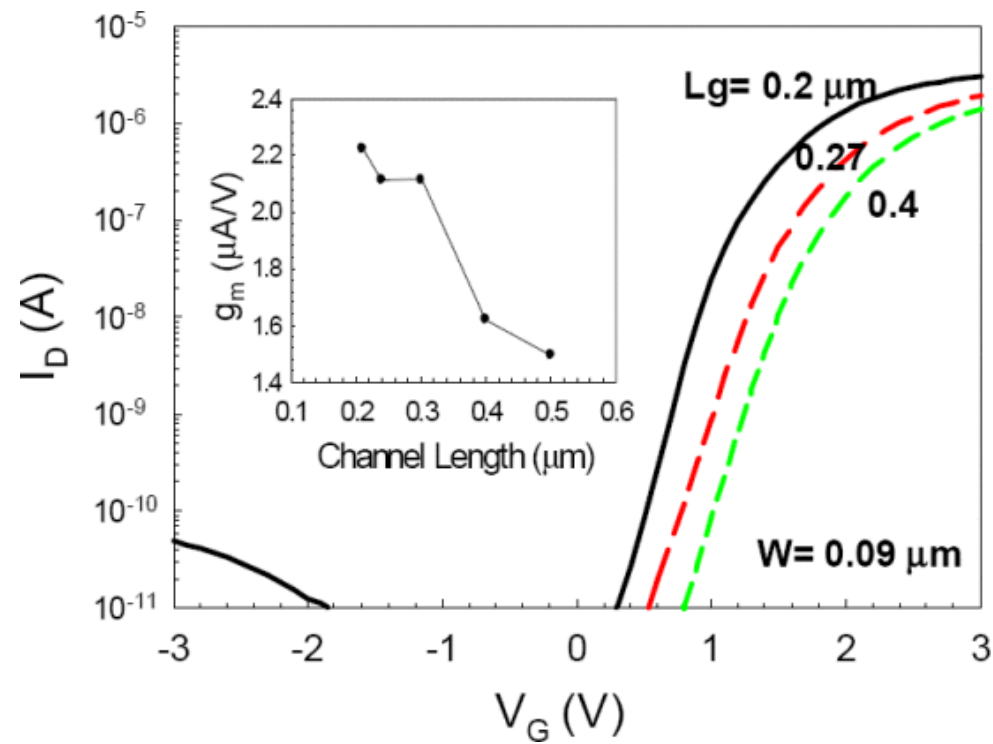


Fig. 5 IV characteristics for various channel length devices. Inset shows the corresponding transconductance (g_m). The channel width is 0.09 μm

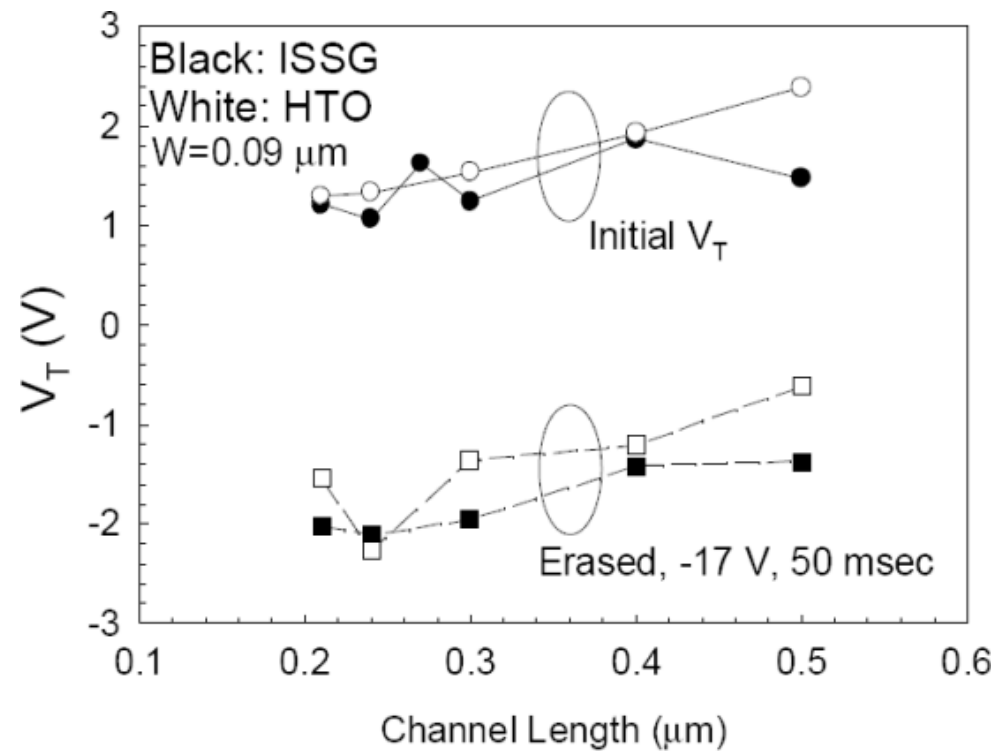


Fig. 6 Short-channel V_T roll-off characteristics for both the initial state and the -FN erased state, with ISSG and HTO top oxide.

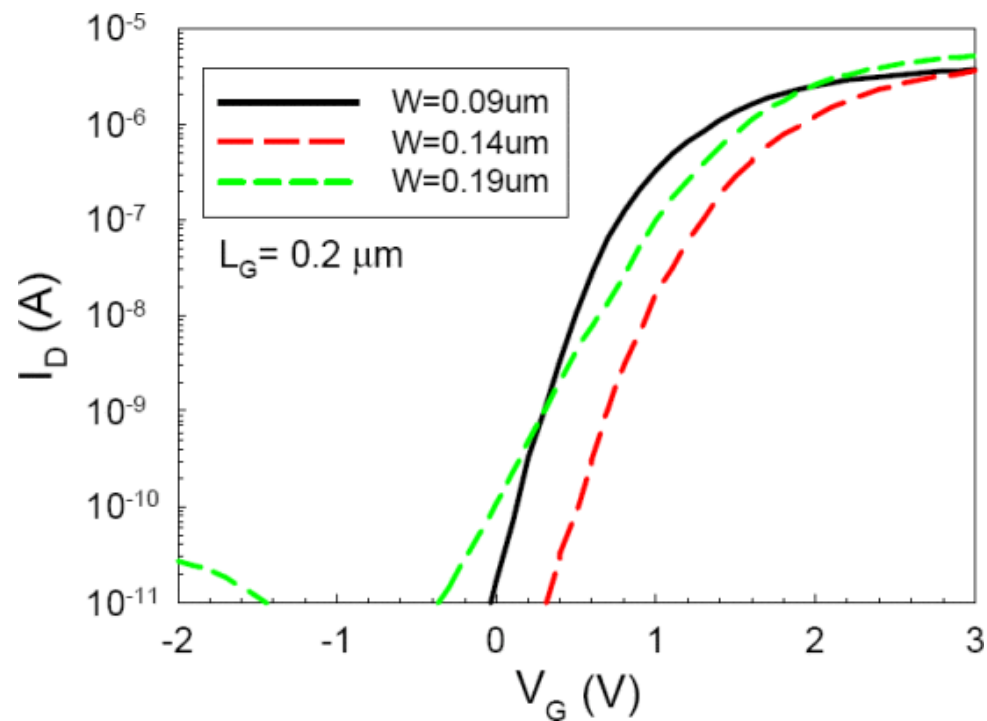


Fig. 7 IV characteristics for various channel width devices. Wider devices show worse subthreshold behavior. The channel length is $0.2 \mu\text{m}$

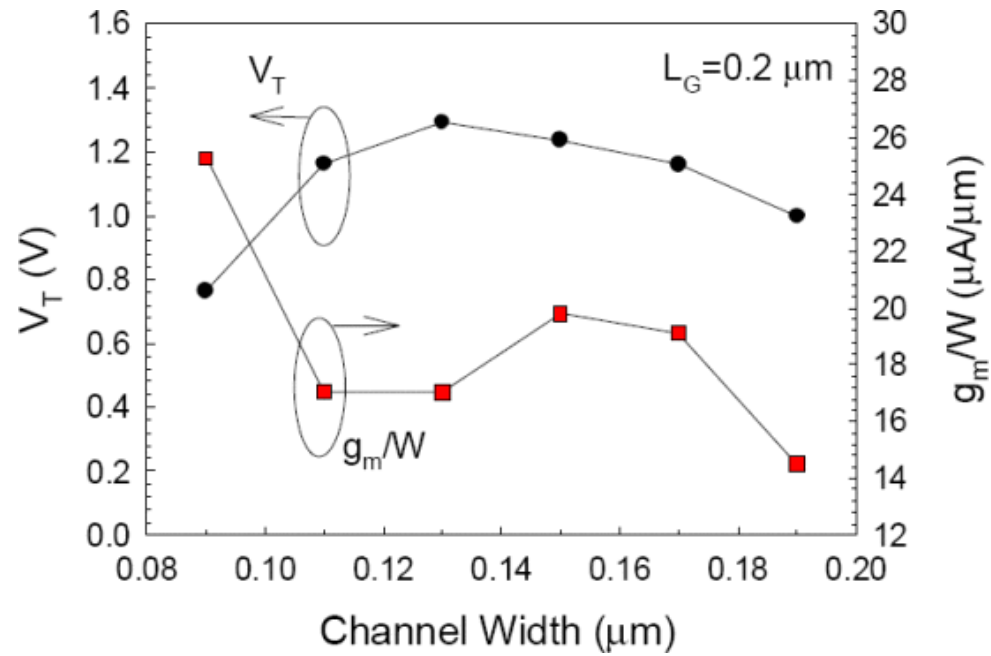


Fig. 8 Narrow-width characteristics. Both normalized transconductance (g_m/W) and V_T are shown. The narrowest device ($W = 0.09 \mu\text{m}$) has the best performance.

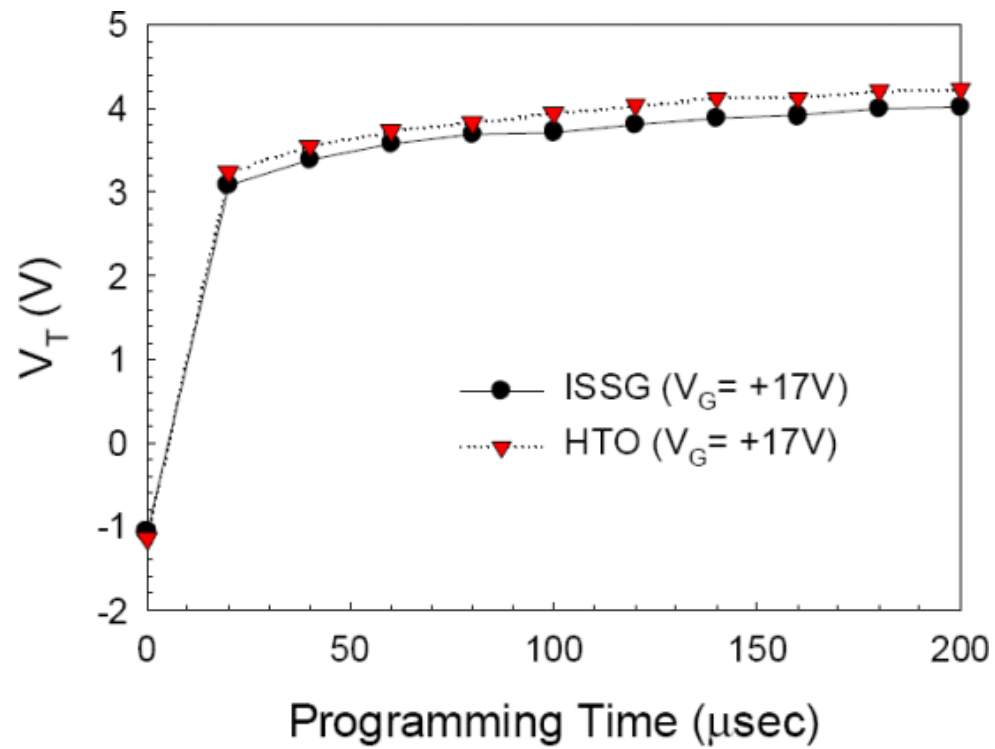


Fig. 9 Comparison of the +FN program characteristics for ISSG and HTO top oxide.

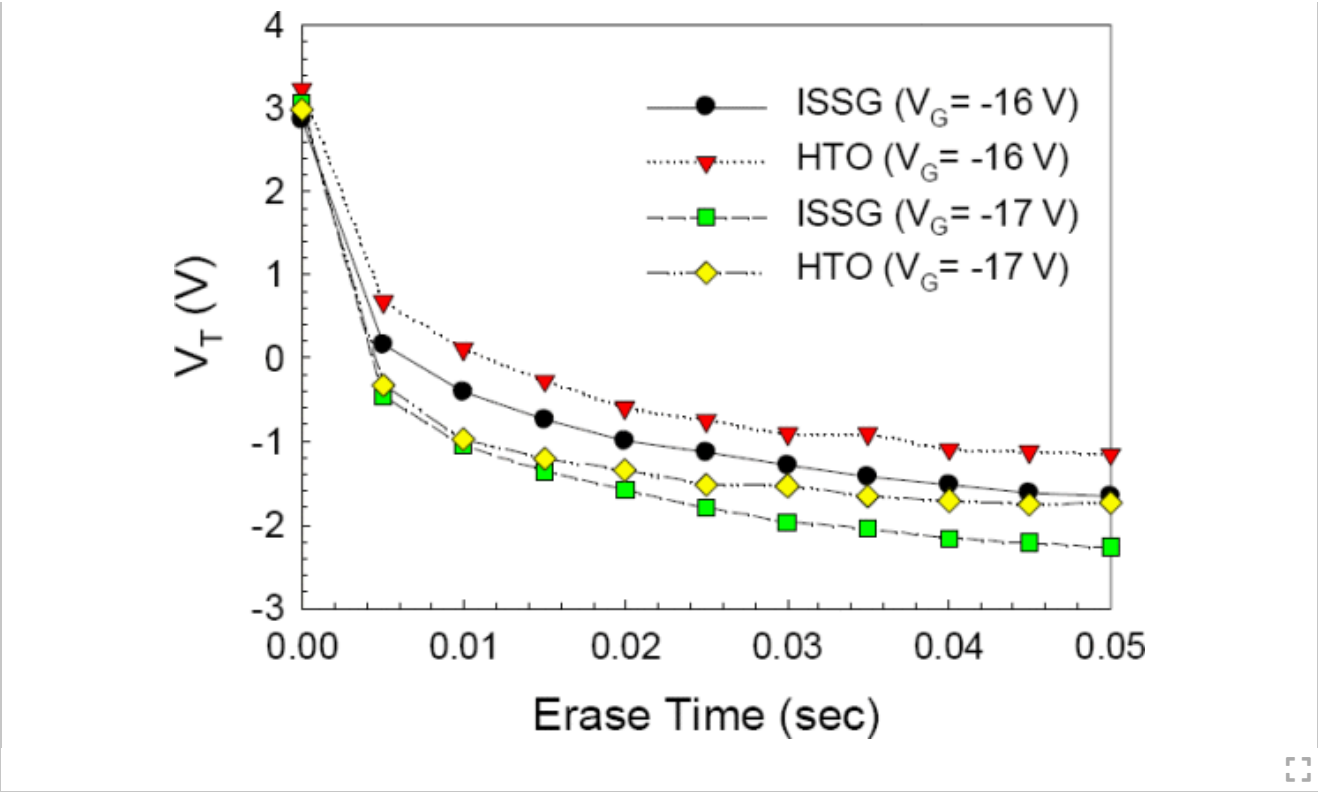


Fig. 10 Comparison of the-FN erase characteristics for ISSG and HTO top oxides

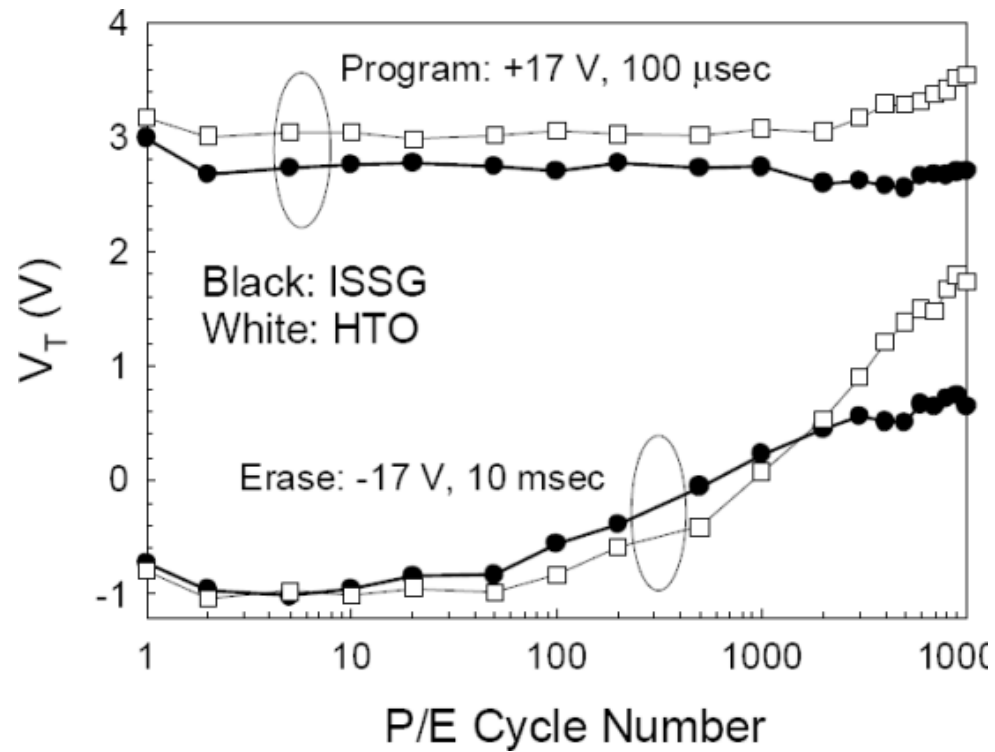


Fig. 11 Dumb-mode P/E cycling endurance (without adjusting the P/E conditions during endurance test). Both ISSG and HTO top oxide are compared.

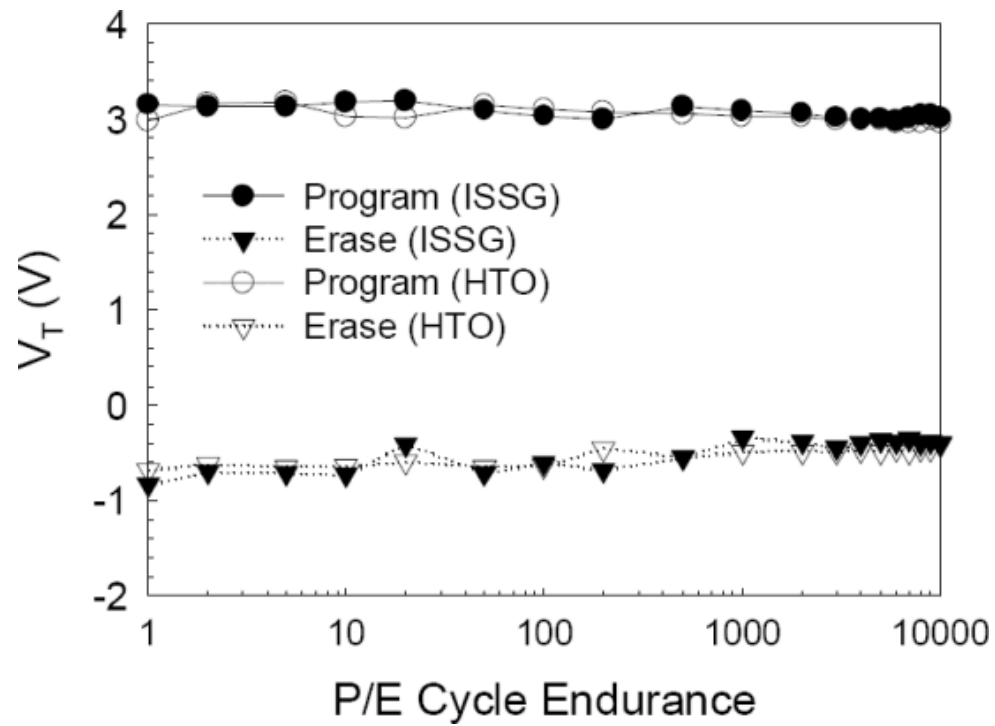


Fig. 12 P/E cycling endurance (with adjusted P/E conditions during endurance test). Both ISSG and HTO top oxide are compared.

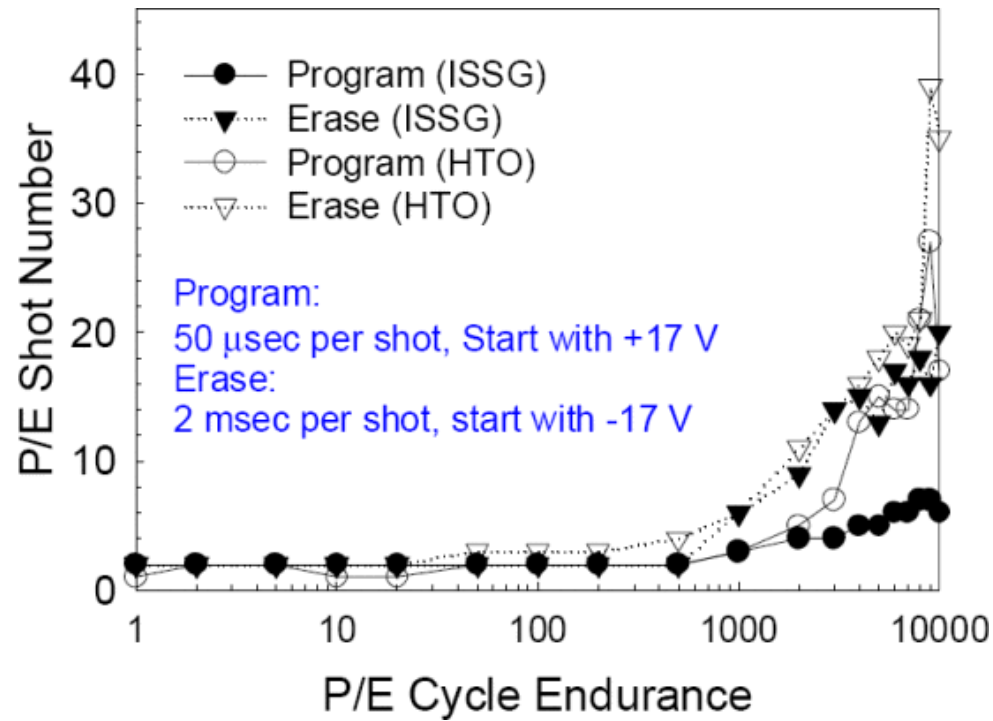


Fig. 13 The required P/E shot number during adjusted P/E tests in Fig. 12. Both ISSG and HTO top oxide are compared.

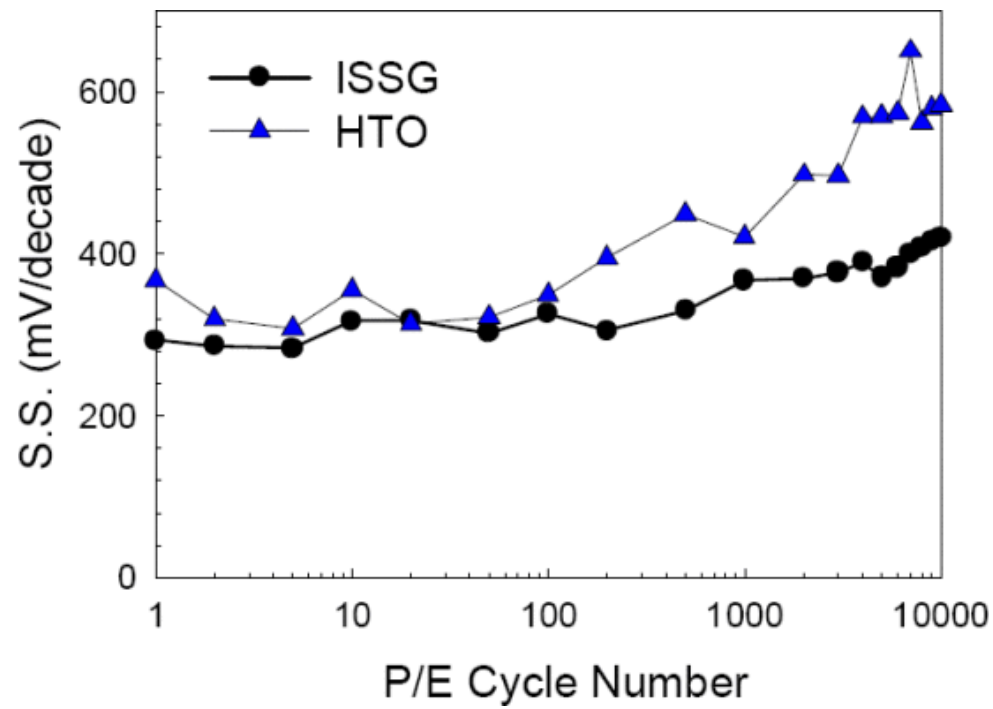


Fig. 14 Subthreshold slope degradation during 10K P/E cycling tests. Both ISSG and HTO top oxide are compared.

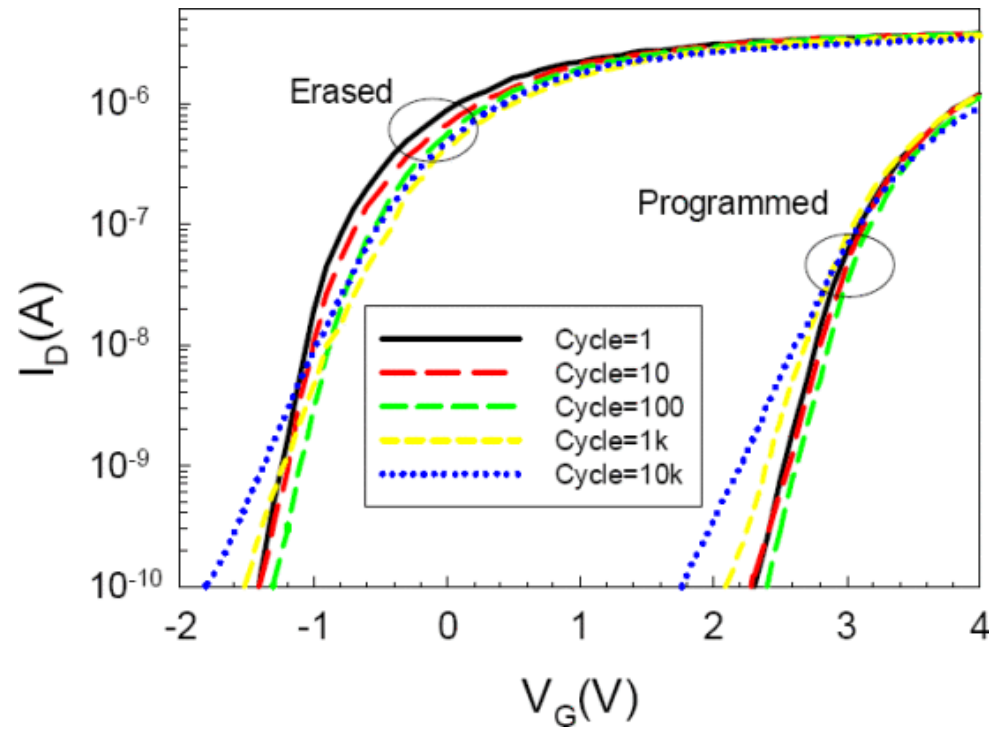


Fig. 15 IV characteristics of ISSG top oxide device during 10K P/E cycling endurance testing.

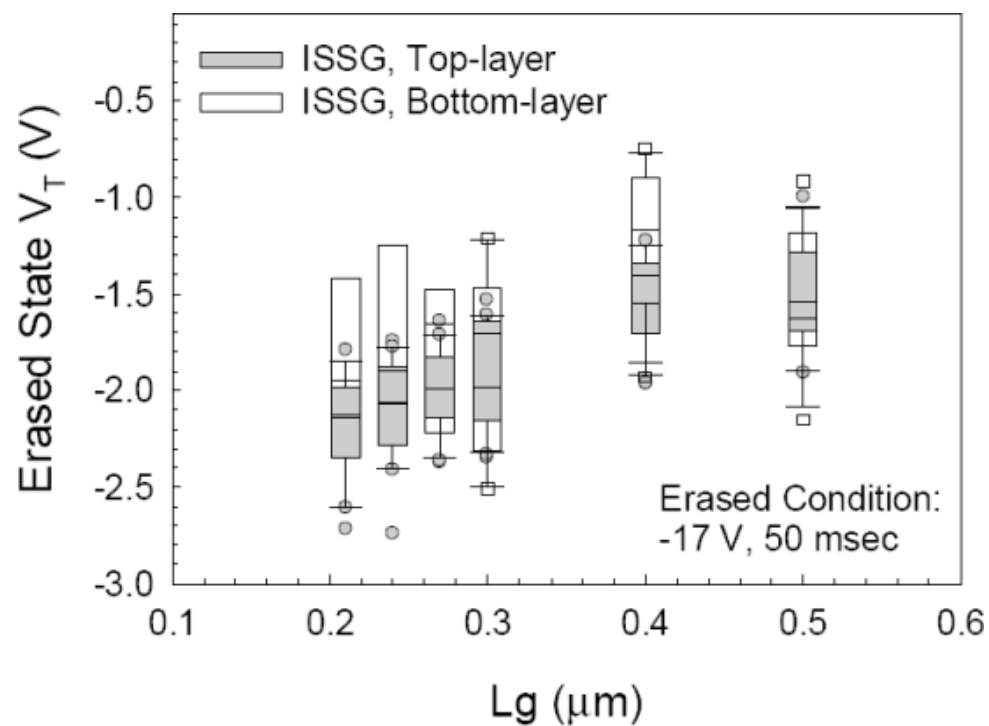


Fig. 16 Short-channel effect comparison of the top and bottom layer TFT NAND devices. A total of 28 identical devices are measured. The devices are in erase saturation state.

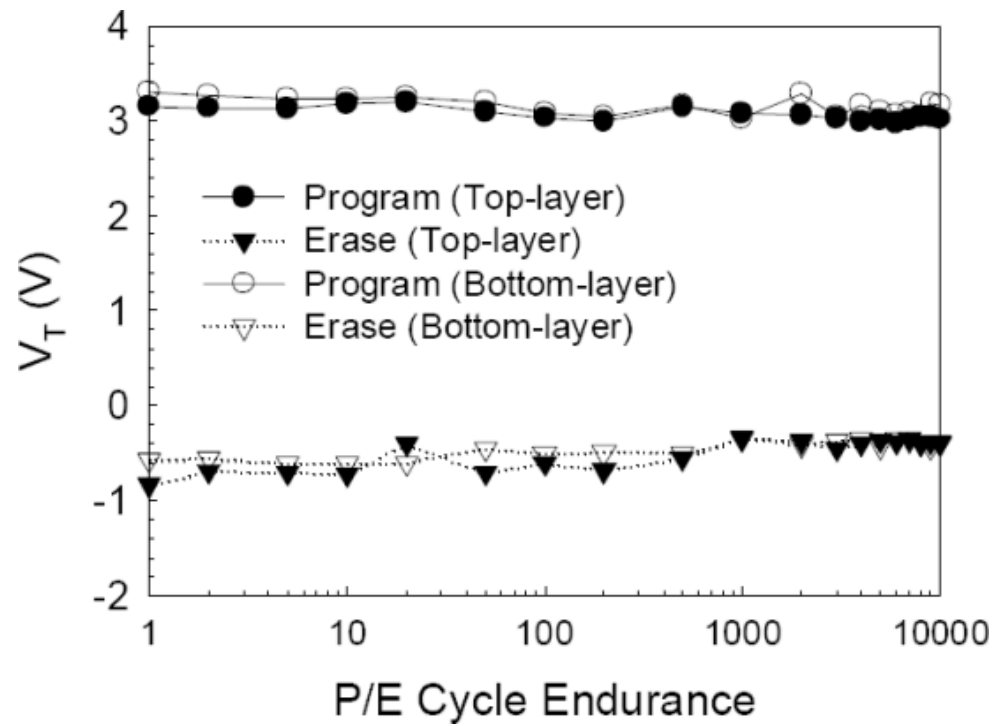


Fig. 17 Adjusted P/E cycling endurance (with adjusted P/E conditions during endurance test). Both bottom and top layers are compared. ISSG top oxide is used.

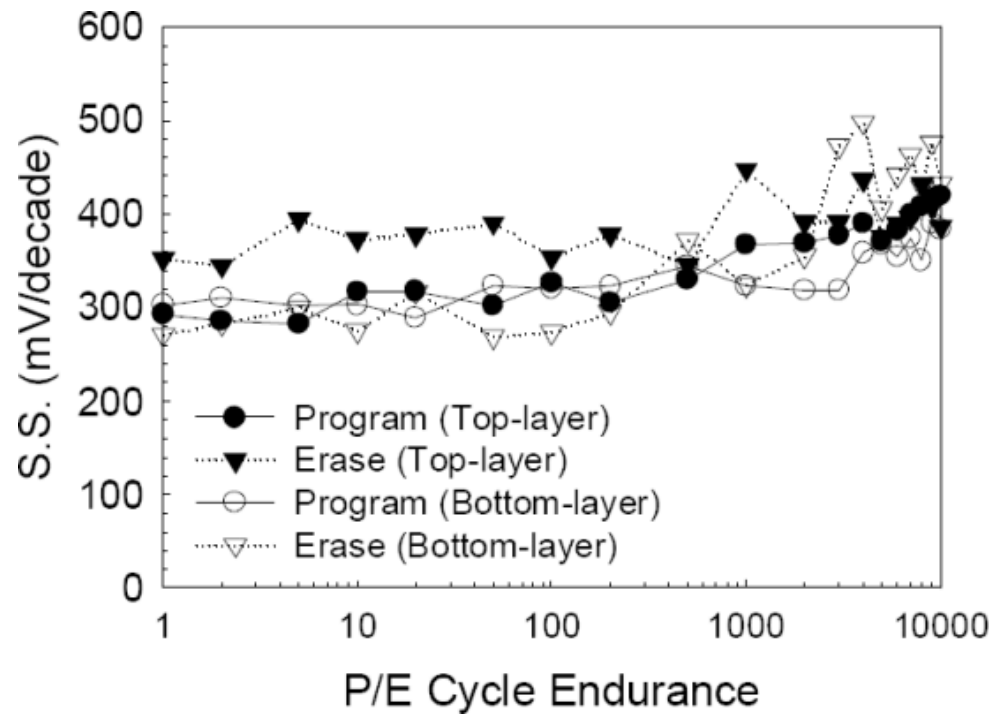


Fig. 18 Subthreshold slope degradation during 10K P/E cycling tests. Both bottom and top layers are compared. ISSG top oxide is used.

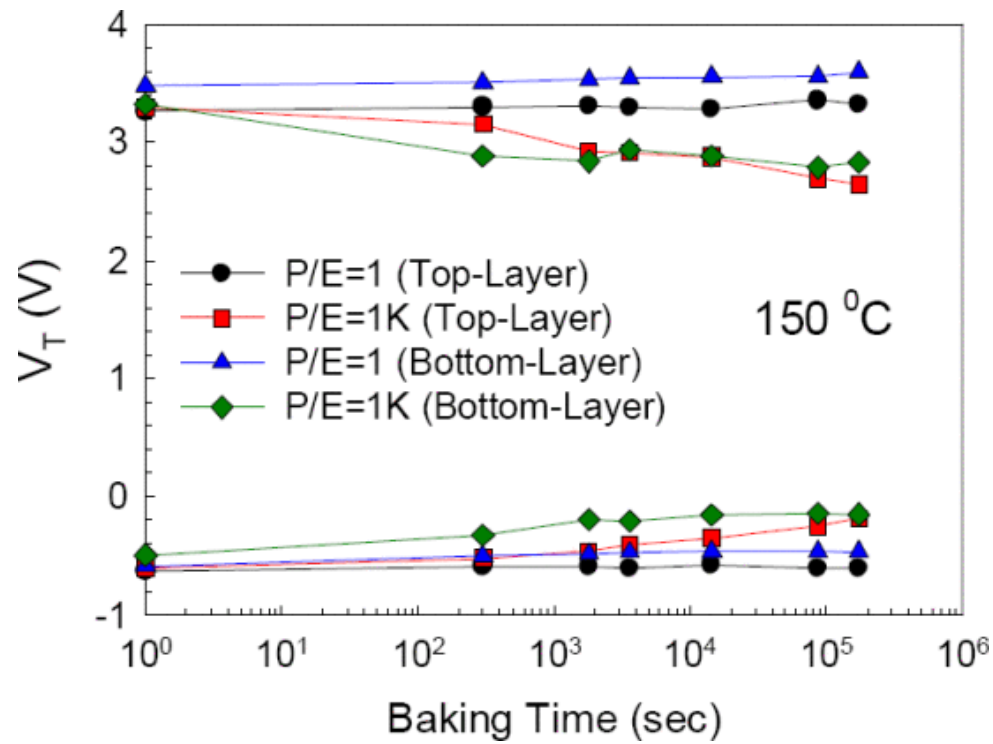


Fig. 19 Retention characteristics. Programmed and erased state with various cycle endurance are compared.

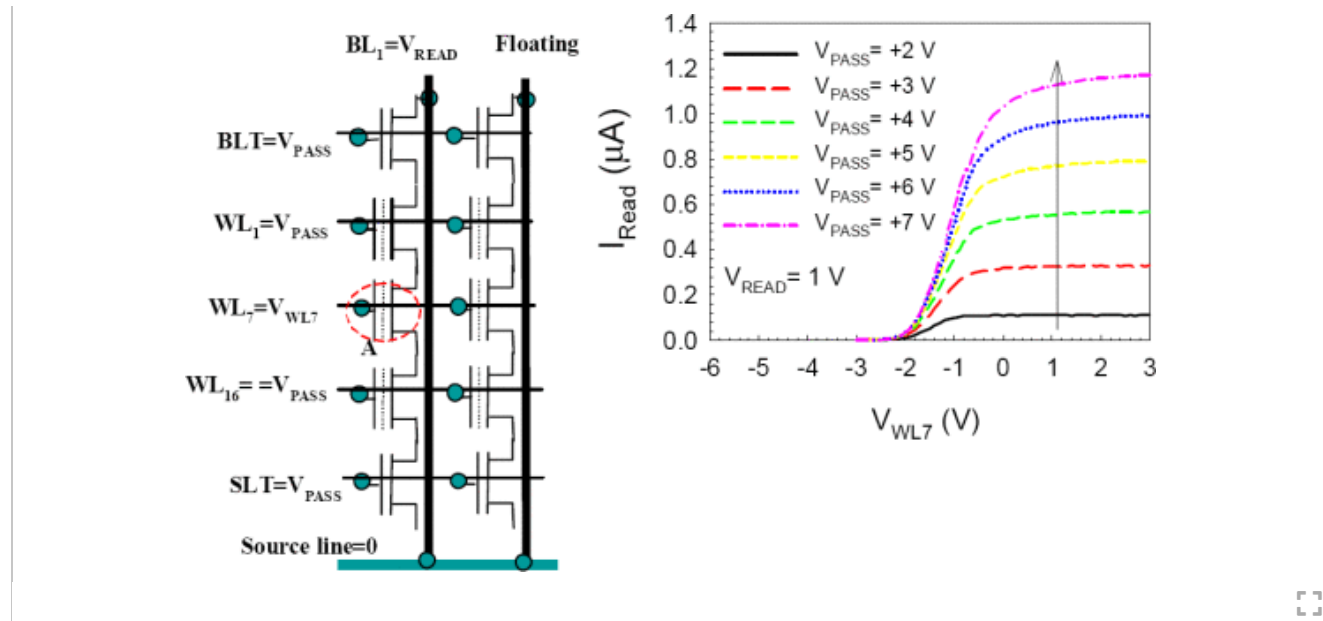


Fig. 20 Read current of a 16 WL's TFT NAND array. The devices are in erased state. For $V_{PASS} = +7$ V, read current can approach $1 \mu\text{A}$, sufficient for NAND Flash design.

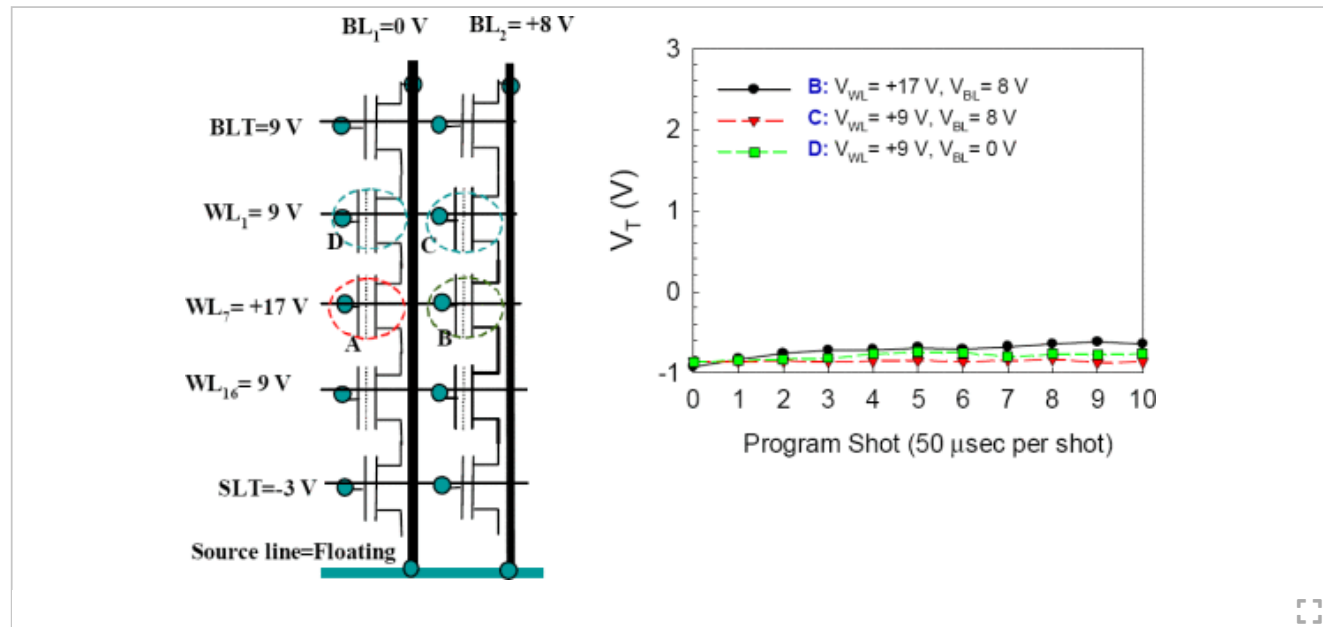
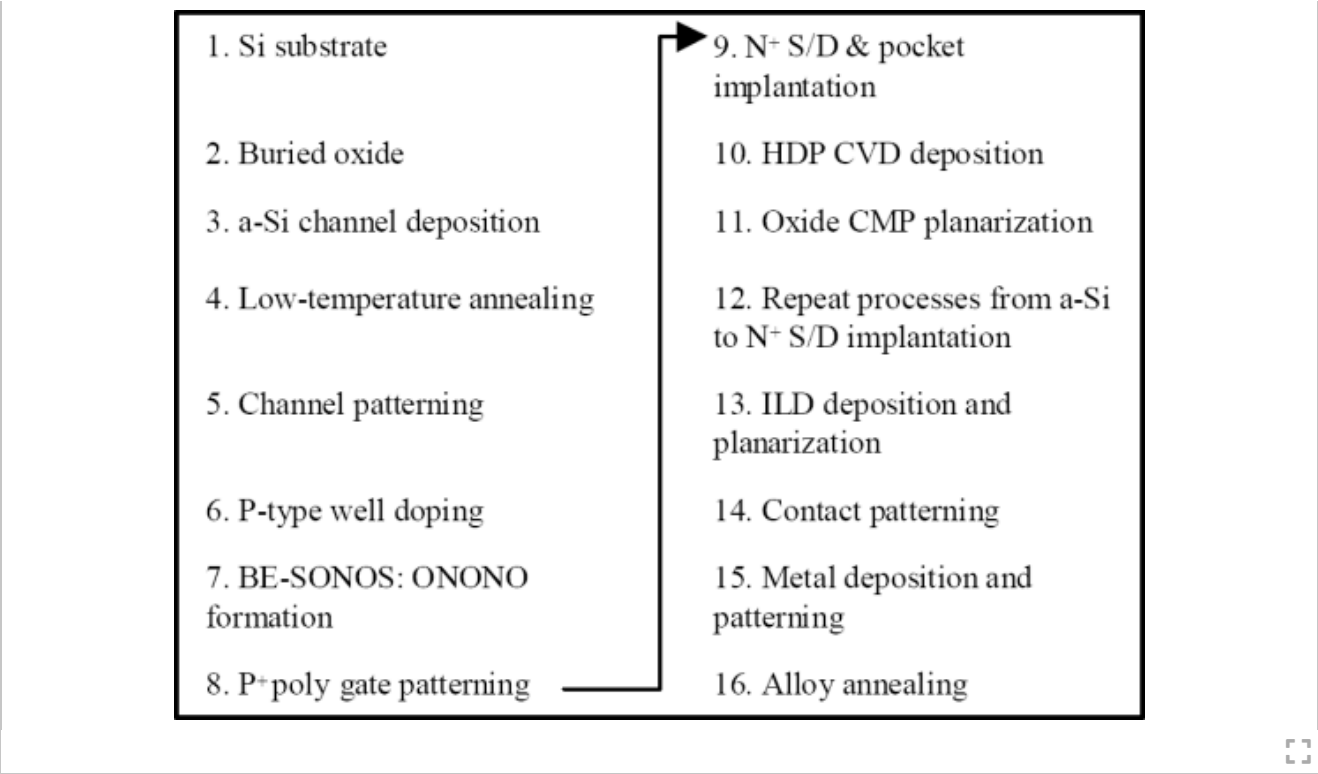


Fig. 21 Program disturb of a TFT NAND array. Cells C and D are under medium field gate disturb, while cell B can be program-inhibited by raising the channel potential.

Table 1 Typical Process Condition of TFT NAND Device.

Poly channel thickness	60 nm
O1	1.5 nm (ISSG)
N1	2 nm (LPCVD)
O2	2 nm (LPCVD)
N2	7 nm (LPCVD)
O3	9 nm (ISSG or HTO)
Poly gate	P ⁺ -poly
Channel length	0.2 μm
Channel width	0.09 μm

Table 2 Fabrication Flow of Multi-Layer TFT NAND Device



ACKNOWLEDGEMENTS

Special thanks to Dr. R.I. Huang, Dr. S.P. Hong, Martin Wu, Dr. Chris Luoh, Terence Hung, C.F. Chen, Bear Wu and Tammy Lin, Technology Development Center in Macronix Inc., for the important cooperation and contributions to this work.

Authors



Figures



References



Citations



Keywords



Metrics



[Back to Results](#)

IEEE Personal Account

CHANGE
USERNAME/PASSWORD

Purchase Details

PAYMENT OPTIONS
VIEW PURCHASED
DOCUMENTS

Profile Information

COMMUNICATIONS
PREFERENCES
PROFESSION AND
EDUCATION
TECHNICAL INTERESTS

Need Help?

US & CANADA: +1 800
678 4333
WORLDWIDE: +1 732
981 0060
CONTACT & SUPPORT

Follow



[About IEEE Xplore](#) | [Contact Us](#) | [Help](#) | [Accessibility](#) | [Terms of Use](#) | [Nondiscrimination Policy](#) | [IEEE Ethics Reporting](#)  | [Sitemap](#) | [IEEE Privacy Policy](#)

A public charity, IEEE is the world's largest technical professional organization dedicated to advancing technology for the benefit of humanity.

© Copyright 2025 IEEE - All rights reserved, including rights for text and data mining and training of artificial intelligence and similar technologies.

Ex. C Abstract

Experimental Characterization of the Vertical Position of the Trapped Charge in Si Nitride-Based Nonvolatile Memory Cells

Publisher: IEEE

Cite This

 PDF

Antonio Arreghini ; Francesco Driussi ; Elisa Vianello ; David Esseni ; Michiel J. van Duuren ; Dusan S. Golubovic

All Authors

42
Cites in
Papers

1
Cites in
Patent

1040
Full
Text Views



Abstract

Document Sections

I. Introduction

II. Experimental Setup and Charge-Centroid-Extraction Procedure

III. Device Description

IV. Carrier-Separation Experiments

V. Position of the Charge Centroid

Show Full Outline ▾

Abstract:

We present a broad set of experiments on silicon nitride-based memories aimed at the investigation of the vertical position of the charge trapped in the nitride layer of silicon-oxide–nitride–oxide-semiconductor (SONOS) memories during program and erase in the tunneling regime. The results obtained for SONOS devices with conventional oxide–nitride–oxide and oxide–nitride–oxide–nitride–oxide gate stacks, as well as with high- κ top dielectric, have been validated by comparing different characterization techniques. It has been shown that, for SONOS cells, the charge centroid is located in the center of the silicon nitride layer, and its position is quite insensitive to the program or erase conditions and to the gate-stack composition.

Published in: IEEE Transactions on Electron Devices (Volume: 55 , Issue: 5, May 2008)

Page(s): 1211 - 1219

DOI: 10.1109/TED.2008.919713

Date of Publication: 31 May 2008 ?

Publisher: IEEE

ISSN Information:

 Authors

 Figures

SECTION I.

References

Citations

Keywords

Metrics

More Like This

Introduction

Further scaling of modern floating-gate (FG) memory cells is hampered by the charge loss due to the stress-induced leakage current (SILC) in cycled cells [1]. In particular, SILC prevents the tunnel-oxide thinning below 8 nm [2] and, moreover, raises statistical-reliability issues for the nowadays very large cell arrays [3], [4]. Furthermore, a strong lateral scaling of nonvolatile-memory (NVM) arrays leads to reliability issues related to the parasitic capacitive coupling between adjacent FG cells [5].

Nitride-based memories, like the silicon-oxide–nitride–oxide–semiconductor (SONOS) devices, are good candidates to replace FG cells in the Flash memory chips of the future generations for high-density and embedded applications, owing to the SILC immunity achieved with localized-charge trapping, that guarantee very good endurance characteristics [6], [7]. Moreover, the relatively low program/erase (P/E) voltages and the excellent compatibility with the standard CMOS process of SONOS memories make them suitable for embedded applications [8]. Finally, the absence of erratic tail bits reduces the statistical-reliability issues in very large memory arrays, and the limited capacitive coupling between adjacent SONOS cells is beneficial for the reliability issues related to the scaling of the spacing between cells.

However, the data retention of SONOS cells programmed and erased by direct tunneling is still a critical issue because of the thin tunnel oxide (≈ 2.5 nm) needed for the tunneling of substrate holes under negative gate voltages [9]. Therefore, in order to optimize the retention characteristics of SONOS memories, an accurate analysis of the trapping properties of the silicon nitride layer that is responsible for the charge storage in the cell is required. In this respect, the knowledge of the location of the charge trapped in the nitride and the type of carriers injected during the P/E operations are not yet complete [10]–[13] despite the importance of these pieces of information in order to understand the charge-loss rates from the nitride toward the gate and substrate and, therefore, to model accurately the retention characteristics.

In this framework, this paper presents an extensive analysis of the vertical position of the charge trapped in the gate dielectric for a wide set of SONOS devices with different oxide–nitride–oxide (ONO) stacks, including samples with nonconventional gate stacks and with a high- κ top dielectric. To accomplish this, we have extensively applied the experimental procedure developed in [14] to extract the trapped-charge centroid, thus fully validating it over several gate stacks for memory cells. Compared with other methods used to probe the charge trapped in the gate stack, our technique is based on the simplest MOSFET electrostatics, and no constraints are imposed to the P/E conditions during its application. In particular, whereas the low-frequency charge pumping (LFCP) [15] can extract the trap profile, but only within a few nanometers inside the ONO

stack, our procedure can reach the deeper locations inside the gate stack, thus considering the effect of the whole trapped charge. However, the centroid alone does not provide the detailed knowledge of the charge profile inside the nitride, but it determines the effectiveness of a given distribution of charges in generating a corresponding V_T shift. Therefore, our technique can be seen as complementary to other methods used to probe the trapped-charge profile (e.g., the LFCP), overcoming some of their limitations in terms of accuracy and applicability.

A broad range of P/E conditions, i.e., different programming/erasing gate voltages (V_G), have been tested for each device, and the experimental results for the charge-centroid location have been confirmed by the comparison with those derived with the LFCP method, thus demonstrating the advantages and limitations of each technique. In a limited number of cases, pMOS SONOS devices were fabricated, and carrier-separation experiments were performed to identify the type of carriers involved in the program/erase operation.

SECTION II.

Experimental Setup and Charge-Centroid-Extraction Procedure

[Fig. 1\(a\)](#) shows the vertical structure of a conventional SONOS device. The charge that is responsible for the threshold-voltage shift (ΔV_T) is supposed to be stored in the silicon nitride layer, which is sandwiched between the tunnel and the top oxides (this assumption will be relaxed later). If we let x_C denote the centroid of the trapped charge measured from the top-oxide/nitride interface [see [Fig. 1\(a\)](#)], then the threshold-voltage shift induced in the cell by the charge Q_N (uniformly injected over the device area A) can be calculated as [\[14\]](#), [\[16\]](#)–[\[18\]](#)

$$\begin{aligned}\Delta V_T &= -\frac{Q_N}{A} \cdot \frac{t_{\text{TOP}}}{\epsilon_{\text{OX}}} - \frac{1}{\epsilon_N} \int_0^{t_N} x \cdot q(x) dx \\ &= -\frac{Q_N}{A} \left(\frac{t_{\text{TOP}}}{\epsilon_{\text{OX}}} + \frac{x_C}{\epsilon_N} \right)\end{aligned}\tag{1}$$

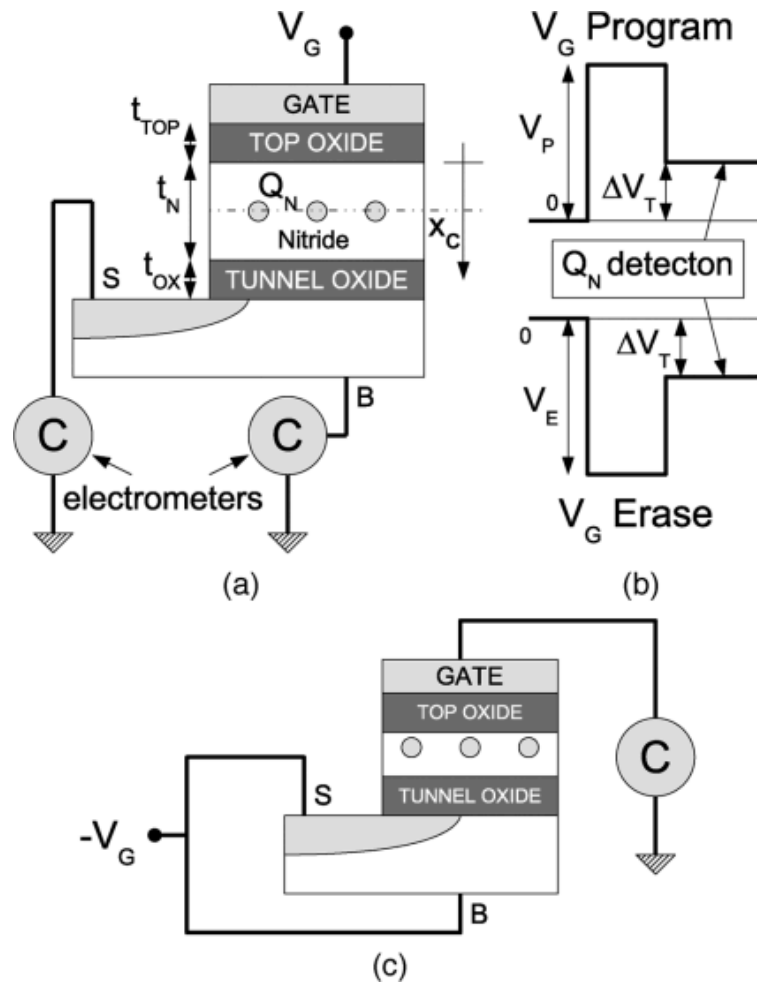
[View Source](#)

where $q(x)$ is the charge density along the nitride depth x , ϵ_{OX} and ϵ_N are the permittivity of SiO_2 and Si_3N_4 , respectively, and t_{TOP} and t_N are the top oxide and nitride thicknesses, respectively. Therefore, if ΔV_T and Q_N are known, the centroid of the charge trapped in the nitride can be univocally determined as

$$x_C = -\epsilon_N \left(\frac{\Delta V_T \cdot A}{Q_N} + \frac{t_{\text{TOP}}}{\epsilon_{OX}} \right). \tag{2}$$

[View Source](#)

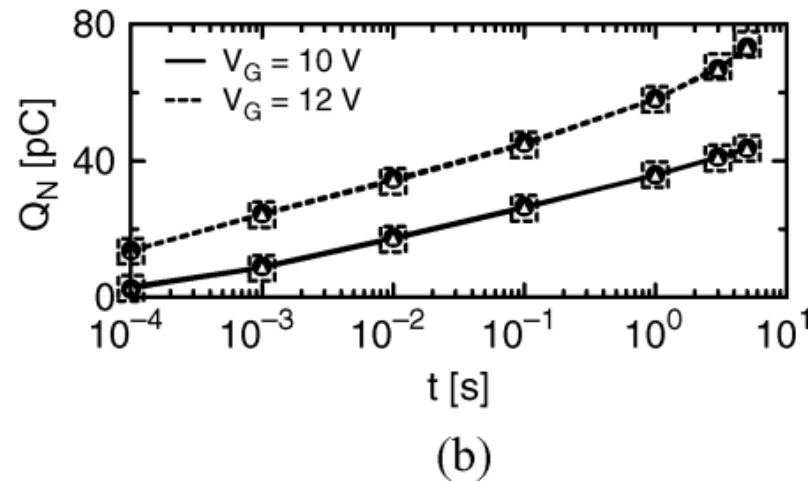
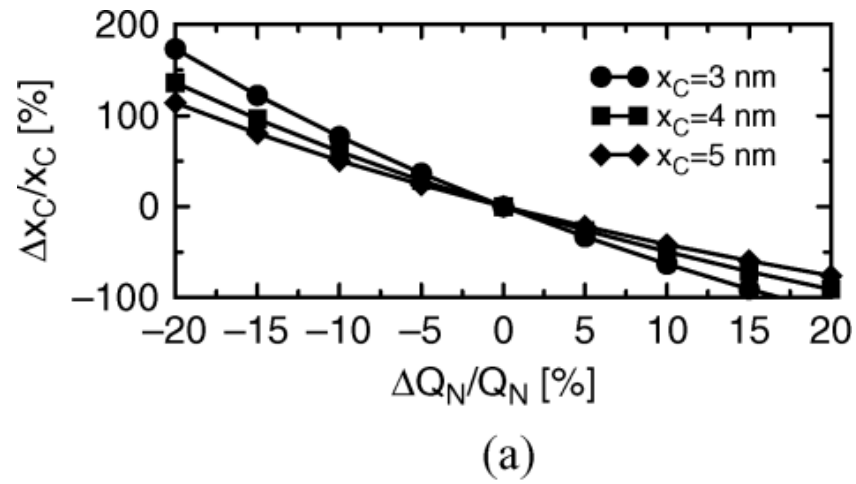


**Fig. 1.**

(a) Measurement setup for carrier separation and charge-centroid extraction. Q_N is the trapped charge, x_C is the centroid position, and t_{TOP} , t_N , and t_{OX} are the thicknesses of the top oxide, nitride layer, and tunnel oxide, respectively. (b) Waveforms for the P/E pulses showing the three-level shape used to obtain a null contribution to the overall current integral of the displacement charge in the semiconductor. (c) Measurement setup for the charge-centroid extraction. By measuring Q_N at the gate while pulsing with a reversed waveform with respect to (b) at the source and bulk terminals, no compensation of possible leakage currents is needed.

Hence, in order to determine x_C by using (2), we have developed a well-controlled experimental setup to directly measure both the programming (erasing) charge Q_N and the corresponding threshold-voltage shift $\Delta V_T = (V_{T\text{ final}} - V_{T\text{ initial}})$ produced by the program (erase) performed in tunneling conditions. For the sake of clarity, it is worth noting that we define program (erase) as the application of gate pulses with positive (negative) V_G , which increase (decrease) the threshold voltage V_T , regardless of the type of the device under test (nMOS or pMOS). In the experimental setup shown in Fig. 1(a), we connected two accurately zeroed electrometers at the source/drain and bulk terminals of the SONOS device. In this way, when an nMOS SONOS cell undergoes a P/E cycle, we can separately measure the contributions to the programming (erasing) charge Q_N of both the electrons detected at source/drain and the holes collected at the bulk, thus performing a carrier-separation experiment.

As for the x_C extraction, an analysis of (2) shows that, while the errors in the detection of ΔV_T are not so critical, x_C is largely sensitive to possible errors on the measured Q_N . With this regard, Fig. 2(a) shows, for a fixed actual centroid position (different symbols), the relative error in the extracted x_C induced by an error in the detection of the charge; we clearly see that even a relatively small error in Q_N can lead to an incorrect estimation of the charge centroid. Therefore, in order to minimize the charge errors, the leakage current through the source-to-bulk junction has been compensated for, first, by measuring its spurious contribution to the charge during a pulse train with a reduced V_G value that prevents the injection into the gate stack and, then, by subtracting the obtained value from the Q_N measured during the P/E pulse with large V_G .

**Fig. 2.**

(a) Sensitivity of the charge centroid on the possible error in the trapped-charge detection. Note that small errors in the measured Q_N lead to large errors in the estimation of x_C . (b) Measured Q_N as a function of the programming pulsewidth for three different devices (circles, squares, and triangles). Note the good repeatability of the measurements obtained with our setup.

Alternatively, because the capacitor is a two-terminal device, the injected charge can be measured also by reversing the roles of the device contacts, hence, through an electrometer connected to the gate, while pulsing on the bulk and source terminals [Fig. 1(c); source and bulk are short-circuited]. In this configuration, a $-V_G$ value is applied during the pulse. In such a way, although carrier separation is not possible, no leakage compensation is needed on the gate terminal (where the charge is measured). Therefore, for the purpose of extracting the charge centroid, the measurement of Q_N by means of the experimental setup in Fig. 1(c) is highly preferable.

Moreover, in order to avoid the spurious contribution to Q_N due to a net, which is the nonnull-time integral of the displacement current, we forced the same $(V_G - V_T)$, hence the same electric field E_{OX} at the Si/SiO₂ interface, which is before and after the program/erase pulse. This was obtained by applying to the gate three-level pulses, as shown in Fig. 1(b), where the final base level of the pulse is displaced by ΔV_T with respect to the initial base level. This approach requires the knowledge in advance of the ΔV_T induced by each program (erase) pulse, and therefore, the P/E characteristics have been preliminarily measured. Afterwards, the measured ΔV_T 's have been used to set the pulse shape and measure the Q_N , as shown in Fig. 1(c). Because the programming and the erasing are performed by applying a sequence of several P/E pulses, for each measurement, we monitored both the cumulative charge Q_N involved in the whole P/E sequence and the partial charge q_N injected during each individual P/E pulse.

As shown in Fig. 2(b), by using this measurement scheme on a given SONOS device, we could achieve a very good reproducibility of the measured Q_N and ΔV_T among different runs and dies, provided that all the P/E characteristics start from the same initial V_T value. An *ad hoc* reset procedure of the device consisting of several mild P/E cycles allowed us to enforce always the same starting V_T value within tens of millivolts. The possible degradation produced by the P/E cycling has been monitored by means of charge-pumping measurements, and it was found to be negligible in the range of relatively few P/E cycles used for our characterization experiments.

SECTION III.

Device Description

The devices used in this paper are nMOS and pMOS SONOS capacitors with a surrounding source diffusion and a gate area of $A = 3100 \mu\text{m}^2$. For both nMOS and pMOS devices, n+ and p+ polygates are available, and the

substrate doping is approximately 10^{17} cm^{-3} . The vertical dimensions of the ONO gate stack have been extracted by means of transmission electron microscopy (TEM), ellipsometry, and capacitance–voltage (C – V) measurements. The nominal thickness of the thermal tunnel oxide is $t_{\text{OX}} = 2.2 \text{ nm}$ for all SONOS cells, whereas two thicknesses are available for both the dichlorosilane Si_3N_4 layer ($t_{\text{N}} = 3.3$ and 5.7 nm , respectively) and the high-temperature-oxide (HTO) top oxide [$t_{\text{TOP}} = 8.9$ and 5.5 nm , respectively; Fig. 3, device A].

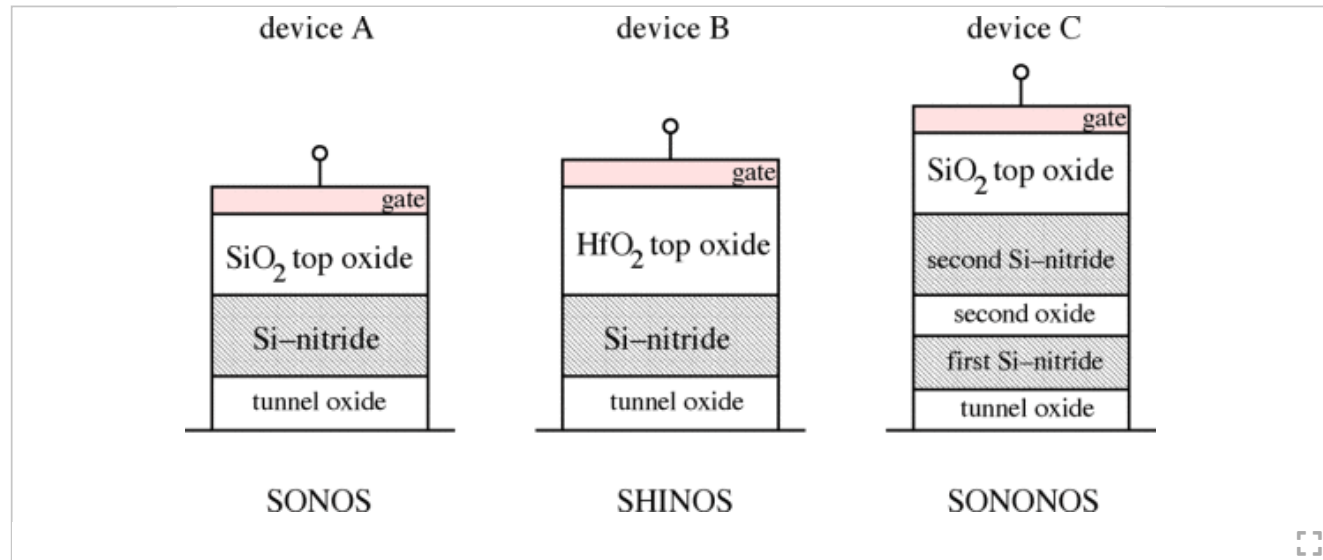


Fig. 3.

Gate-stack composition of the devices used in this paper.

A gate stack featuring a 11.8-nm-thick HfO_2 layer as top dielectric, a 5.7-nm-thick Si nitride, and a 2.2-nm-thick tunnel oxide was also available (hereafter denoted as SHINOS; Fig. 3, device B).

Finally, the nonconventional gate-stack structure shown in Fig. 3 (device C, hereafter denoted as SONONOS) has been also used for the experiments related to the charge-centroid determination. This device is composed of two nitride layers sandwiched between the tunnel and top oxides and separated by an additional SiO_2 layer (second oxide) formed by the reoxidation of part of the first silicon nitride layer (the one on top of the tunnel oxide). In this device, the thicknesses extracted by means of TEM and C – V measurements are 1.3, 2.8, 1.5, 5.0,

and 7.2 nm for the tunnel oxide, the first nitride, the second oxide, the second nitride, and the top oxide, respectively ([Fig. 3](#), device C). All the measured devices are free of the passivation issue discussed in [\[19\]](#).

SECTION IV.

Carrier-Separation Experiments

Carrier-separation experiments are suitable to investigate quantitatively the contribution of electrons and holes during the program and erase operations. Previous experiments showed that, during program, the charge Q_N is almost entirely due to electrons [\[11\]](#), [\[13\]](#), [\[14\]](#), [\[17\]](#), whereas during erase, the contribution of holes can be significant [\[9\]](#), [\[14\]](#). In this regard, we first successfully verified with the experimental setup in [Fig. 1\(a\)](#) that Q_N is due to electrons during program [\[14\]](#). In analogy with the case of quantum-yield experiments, the hole recombination with electrons in the inversion layer is a very unlikely process [\[20\]](#), [\[21\]](#), and, thus, we conclude in agreement with Bachhofer et al. [\[11\]](#), Wang and White [\[13\]](#), and Libsch and White [\[17\]](#) that the hole detrapping is negligible during this phase.

In the case of erasing, because our nMOS SONOS test structures are rather large ($20 \times 155 \mu\text{m}$), the electrons that are possibly detrapped from the nitride must diffuse for several micrometers in the substrate before they can reach the source contact; thus, the recombination with substrate holes can take place, and the carrier-separation technique may underestimate the contribution to the erase of the detrapped electrons. The experiments on nMOS devices revealed that the measured Q_N during erase is mainly due to holes [\[14\]](#); hence, either the electron detrapping is very small or it is masked by the recombination in the p -substrate.

The accuracy of the carrier-separation experiments for determining the contribution of electrons and holes (Q_e and Q_h , respectively) to the erase is significantly improved if we use pMOS SONOS capacitors. In fact, under negative V_G , the holes are supplied by the inversion layer and are measured at the source, whereas the detrapped electrons go through the very thin hole inversion layer and, then, are collected at the bulk terminal of the n -type substrate [\[20\]](#), [\[21\]](#) so that the probability of recombination with holes is drastically reduced with respect to the experiment of the erase in nMOS SONOS.

[Fig. 4](#) shows the carrier-separation measurements performed during the erase transient in pMOS SONOS devices, which, according to our definition, correspond to a reduction of the threshold voltage, i.e., to negative

ΔV_T values. We see how the hole contribution Q_h to Q_N depends on the initial programmed state V_{TP} . In particular, for $V_{TP} = -0.6$ V, we observe a reduced contribution of holes, hence a large contribution of electrons detrapped from the nitride layer and a strong ΔV_T dependence of the (Q_h/Q_N) ratio. At higher V_{TP} , the relative contribution of holes remains constant versus ΔV_T , regardless of the erasing gate voltage V_G . Moreover, if we replot Q_h as in Fig. 5, we see how the hole component directly depends on the initial gate overdrive voltage $(V_G - V_{TP})$. This latter result can be justified by considering that both the tunneling rate and the hole-inversion-layer density depend on the $(V_G - V_{TP})$ value.

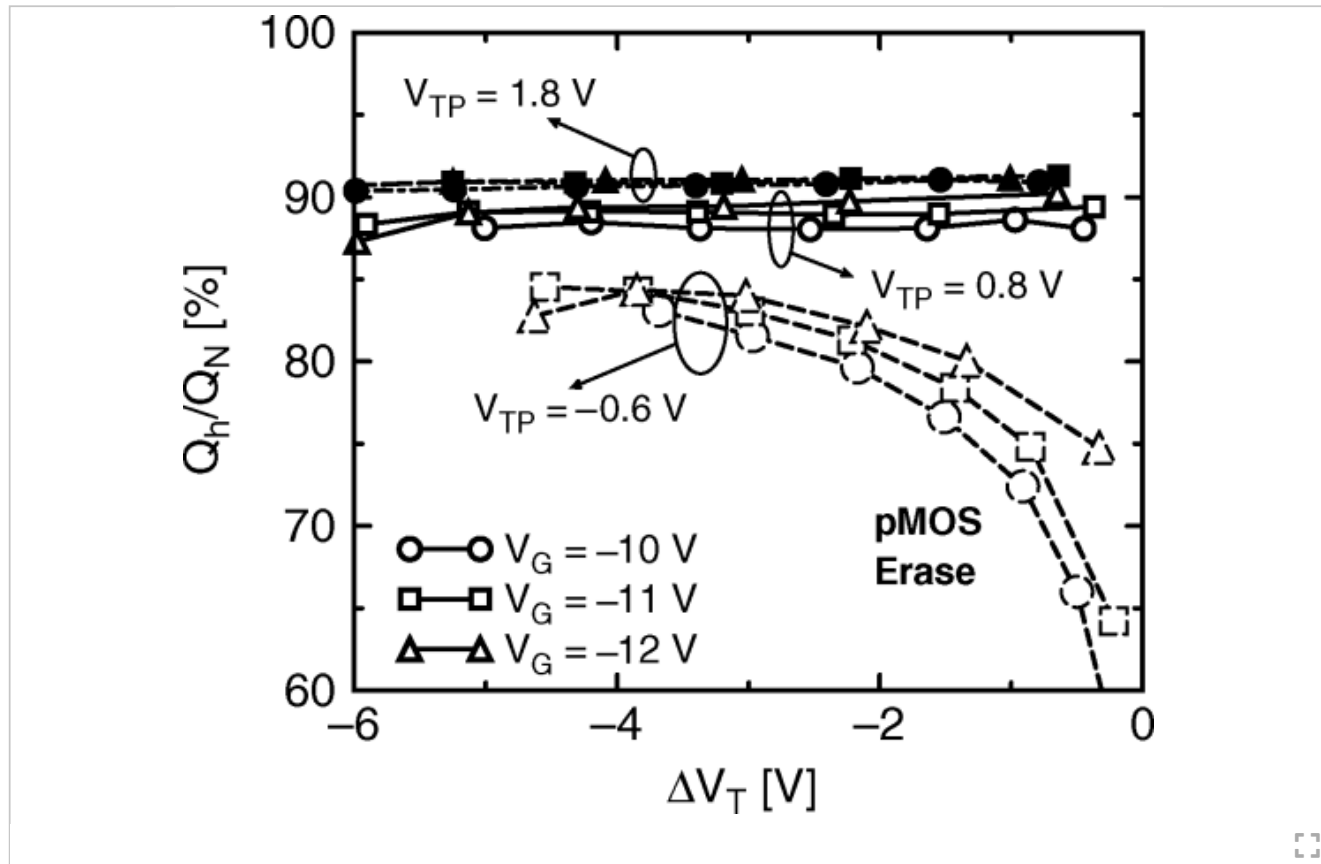
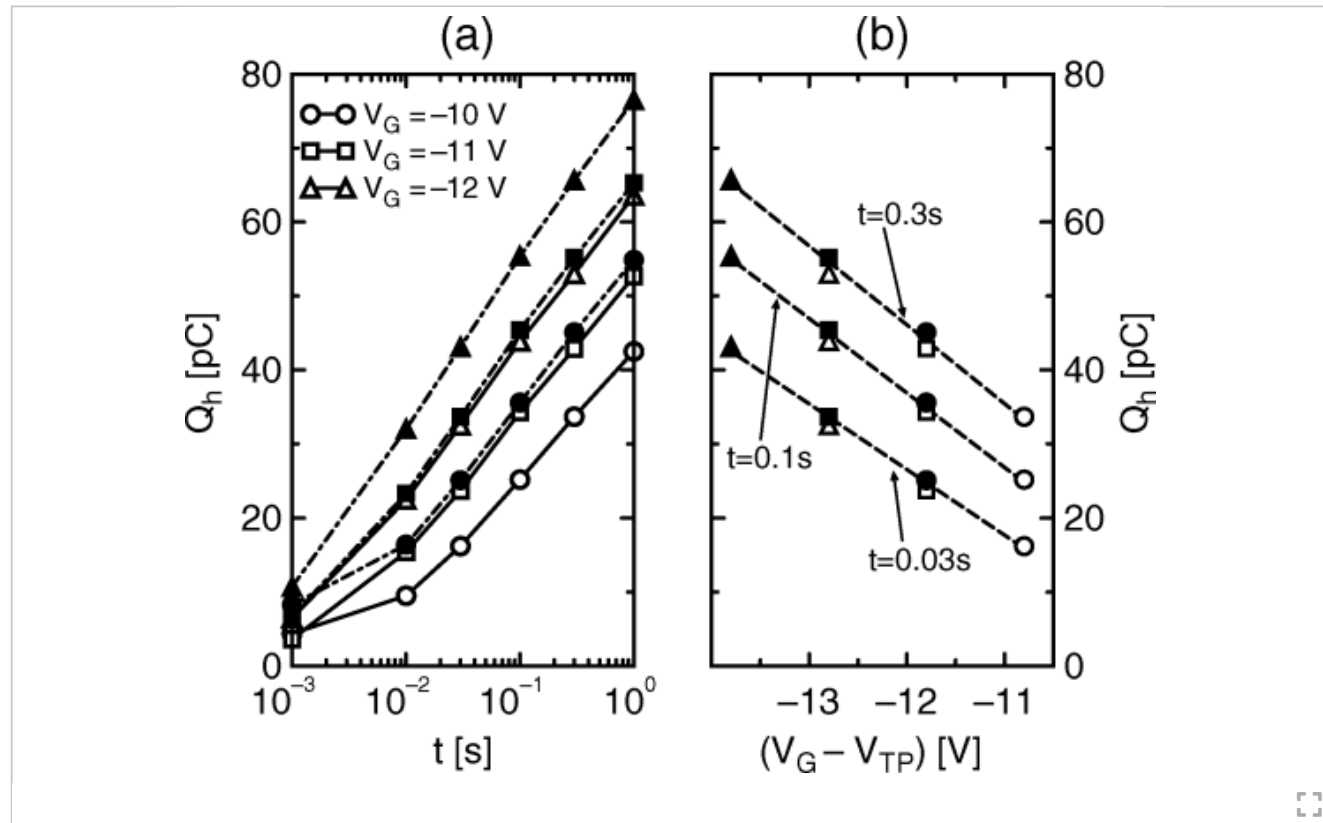


Fig. 4.

Fraction of the charge due to holes measured during the pMOS SONOS erase as a function of ΔV_T for different values of V_G and of the initial threshold voltage V_{TP} . Note the decreasing contribution of holes for decreasing V_{TP} . $Q_h, Q_e > 0$ during erase.

**Fig. 5.**

Hole contribution to pMOS SONOS erase as a function of the (a) erasing time and (b) $(V_G - V_T)$. Open symbols for $V_{TP} = 0.8$ V and filled symbols for $V_{TP} = 1.8$ V.

SECTION V.

Position of the Charge Centroid

After the identification of the different roles played by electrons and holes during P/E, we focused our attention on the extraction of the charge centroid x_C by using (2). In particular, to study the x_C dependence on the P/E

conditions and the device structure, the experimental procedure described in [Section II](#) has been used to measure Q_N and the induced ΔV_T for several V_G values and for devices featuring different ONO gate stacks.

A. Experimental Results for SONOS and SHINOS Devices

[Fig. 6](#) shows the P/E characteristics as functions of the total program (erase) time t of the nMOS SONOS with $t_{OX} = 2.2$ nm, $t_N = 5.7$ nm, and $t_{TOP} = 8.9$ nm (2.2/5.7/8.9). The ΔV_T 's have been evaluated by measuring the shifts of the $C-V$ curve after each P/E pulse. Then, the capacitor was reset to the initial state (by mild P/E cycling), and the ΔV_T data have been used to set the shape of the pulses [see [Fig. 1\(b\)](#)] during which we directly measured the injected Q_N with the experimental setup in [Fig. 1\(c\)](#). Finally, in [Fig. 7\(a\)](#), we have calculated the x_C of the cumulative charge Q_N . Each point of the curve is the centroid of the total charge Q_N injected from the beginning of the P/E characteristic to the time corresponding the ΔV_T value reported in the x -axis in [Fig. 7\(a\)](#).

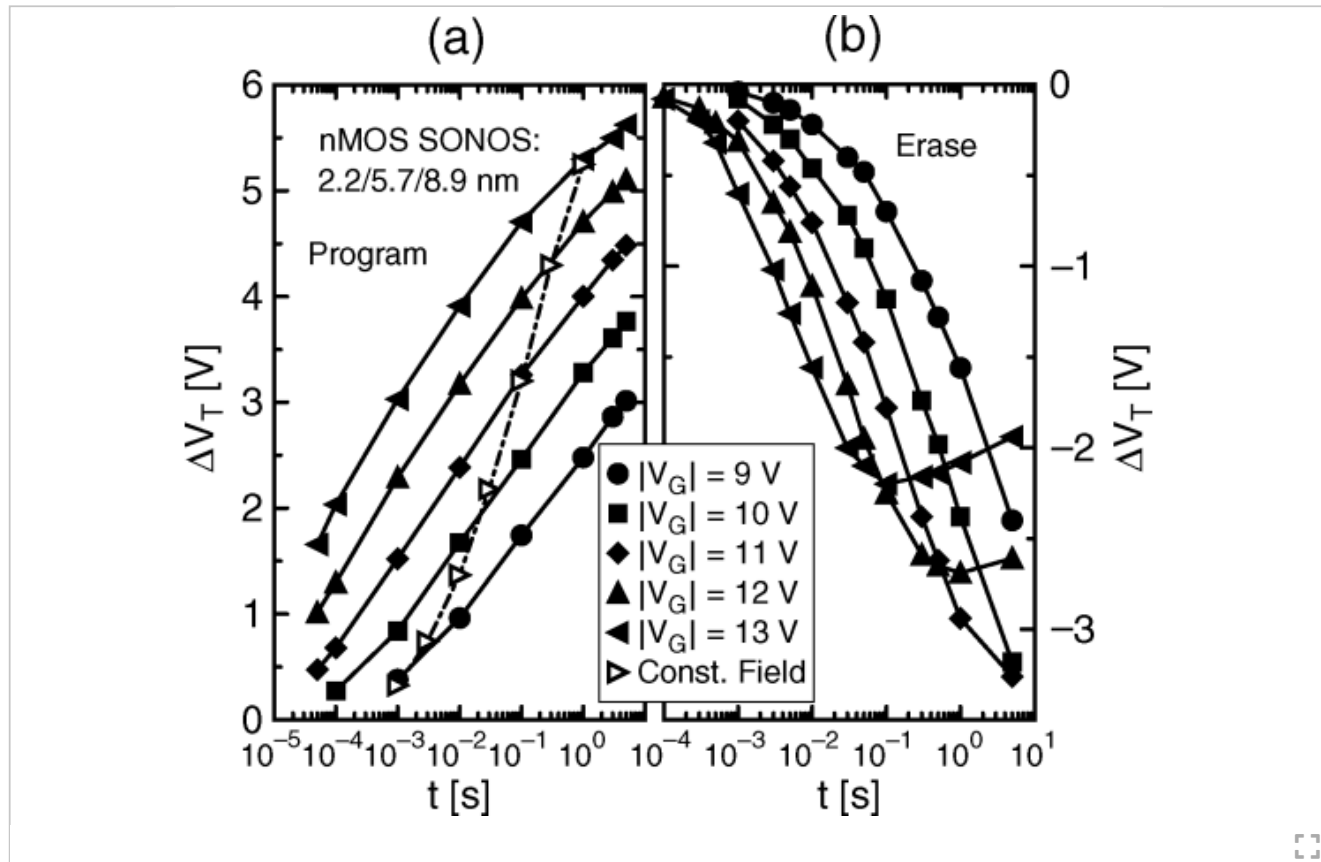


Fig. 6.

(a) Program and (b) erase characteristics of an nMOS SONOS capacitor at constant gate voltage. Open triangles correspond to programming at constant tunnel-oxide field, which is obtained by compensating at each program pulse the ΔV_T produced by the previous pulse.

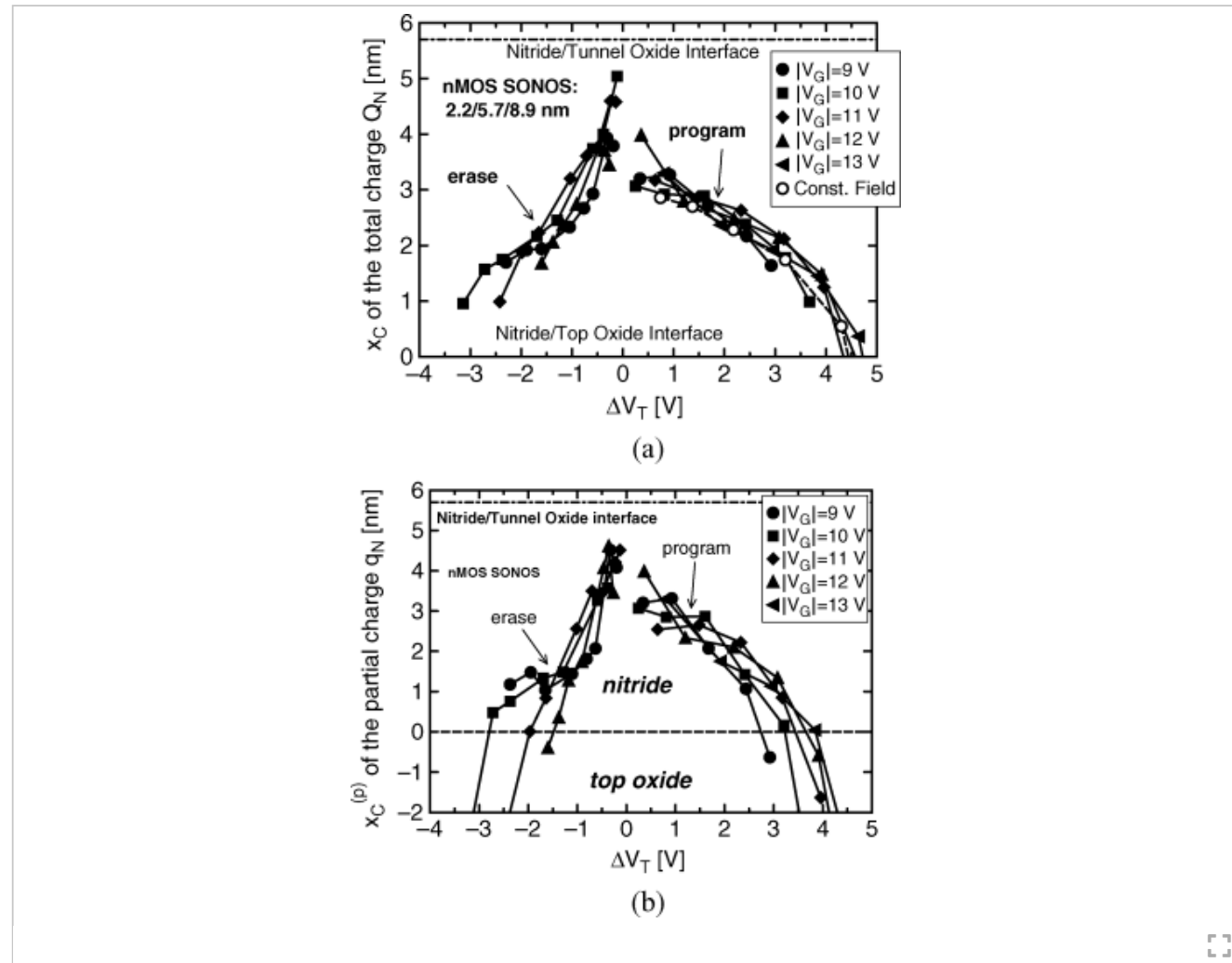


Fig. 7.

(a) Centroid of the cumulative charge trapped in the nitride of the nMOS SONOS capacitor (2.2/5.7/8.9-nm ONO stack) extracted for the program and erase pulses at different V_G 's. Open circles correspond to the programming at constant electric field across the tunnel oxide (V_G is adjusted after each subsequent pulse). (b) Centroid of the partial charge injected into the nitride during each pulse of the programming/erasing of the nMOS SONOS (2.2/5.7/8.9-nm ONO stack). Note how during the last program/erase pulses (large $|\Delta V_T|$) the centroid of the partial charge is located in the top oxide, indicating that part of the injected charge has been lost through the gate.

We observe that the centroids for program and erase are very consistent with each other; at the beginning of the program (erase) phase, x_C is located near the middle of the nitride layer, which is in agreement with the results for HTO top oxides reported in [22]. Then, x_C slowly moves toward the top oxide. Moreover, the data in Fig. 7(a) do not show any dependence on the applied gate voltage V_G . In this respect, in order to further investigate a possible bias dependence of x_C , we have also performed measurements by keeping E_{OX} at the Si/SiO₂ interface approximately constant during program [Figs. 6(a) and 7(a); open symbols] by increasing V_G after each pulse for an amount equal to the corresponding ΔV_T . This, of course, leads to a faster programming [Fig. 6(a); open symbols], but x_C still agrees well with the data obtained at constant V_G [Fig. 7(a)]. Therefore, we can conclude that there is no dependence of x_C on the injection conditions, at least in the bias range explored in this paper.

For $\Delta V_T > 4$ V, the x_C values move toward the top oxide and eventually become negative. The negative x_C , however, is most likely an artifact due to an electron leakage through the top oxide at the largest programming times, hence largest ΔV_T 's. In fact, when the cell is strongly programmed and the nitride is almost completely charged, the electric field across the top oxide is maximum, and the electrons injected in the nitride can tunnel out of the storage layer during the last part of the programming phase. This charge leakage offsets our estimate of Q_N , which hence appears less effective with respect to the measured threshold-voltage shift and results in a smaller value of x_C [see (2)]. In order to more precisely assess the ΔV_T range where the determination of x_C is reliable, in Fig. 7(b), we calculated the centroid $x_C^{(p)}$ of the partial charge packet q_N injected during the last pulse of the P/E sequence used to reach a given ΔV_T value, hence injected between two adjacent points of the curves in Fig. 7(a). The centroid $x_C^{(p)}$ is calculated by inserting in (2) the charge q_N and the threshold-voltage shift corresponding to the single (individual) P/E pulse. It is worth to note how, for $\Delta V_T > 3$ V, $x_C^{(p)}$ is negative, thus indicating that we are measuring a charge packet q_N that is larger than the one needed to justify

the partial threshold-voltage shift obtained during the last P/E pulse. Hence, part of the charge q_N must have been lost through the gate. For $\Delta V_T < 3$ V, on the other hand, the $x_C^{(p)}$ of each partial charge packet is located well inside the nitride, thus indicating that we are not losing charge through the gate and validating the x_C values extracted for the cumulative charge Q_N in this range of ΔV_T values.

Moreover, the analysis of the $x_C^{(p)}$ values shows that the centroid position is independent of the starting V_T , because each point in [Fig. 7\(b\)](#) corresponds to the charge centroid obtained, starting from a different V_T level.

During erase, $x_C^{(p)}$ becomes negative when the electron injection from the gate takes place. In this situation, a steady-state current flows from the substrate to the gate, thus yielding a saturation of the erase characteristics that limit the observable ΔV_T range [[Fig. 6\(b\)](#)]. The charge measured in this condition is largely exceeding the charge responsible for the threshold-voltage shift, and therefore, the extracted x_C becomes unreliable while approaching this region. On the basis of such observations, we estimated that, for the measured device, the x_C values are reliable for $-2 < \Delta V_T < 3$ V and, in this ΔV_T range, the charge centroid stays close to the center of the nitride layer and it only slightly moves toward the top oxide for both program and erase [[14](#)], [[18](#)], [[22](#)].

As stated in [Section IV](#), the carrier-separation experiments showed that, in our nMOS SONOS capacitors, the measured Q_N during program is essentially due to electrons whereas, during erase, it is mainly due to holes. Thus, the net result of a complete P/E cycle is a recombination charge, and we can also use an LFCP technique to detect such a recombined charge Q_N [[15](#)]. [Fig. 8](#) shows the x_C values calculated by using Q_N detected by means of LFCP with different base and top pulse levels. We see that the results agree well with the x_C obtained with the direct measure of Q_N on the same device, except at small ΔV_T , where the LFCP method underestimates Q_N , as shown in [[14](#)], thus resulting in a deviation of x_C toward the tunnel oxide [see ([2](#))].

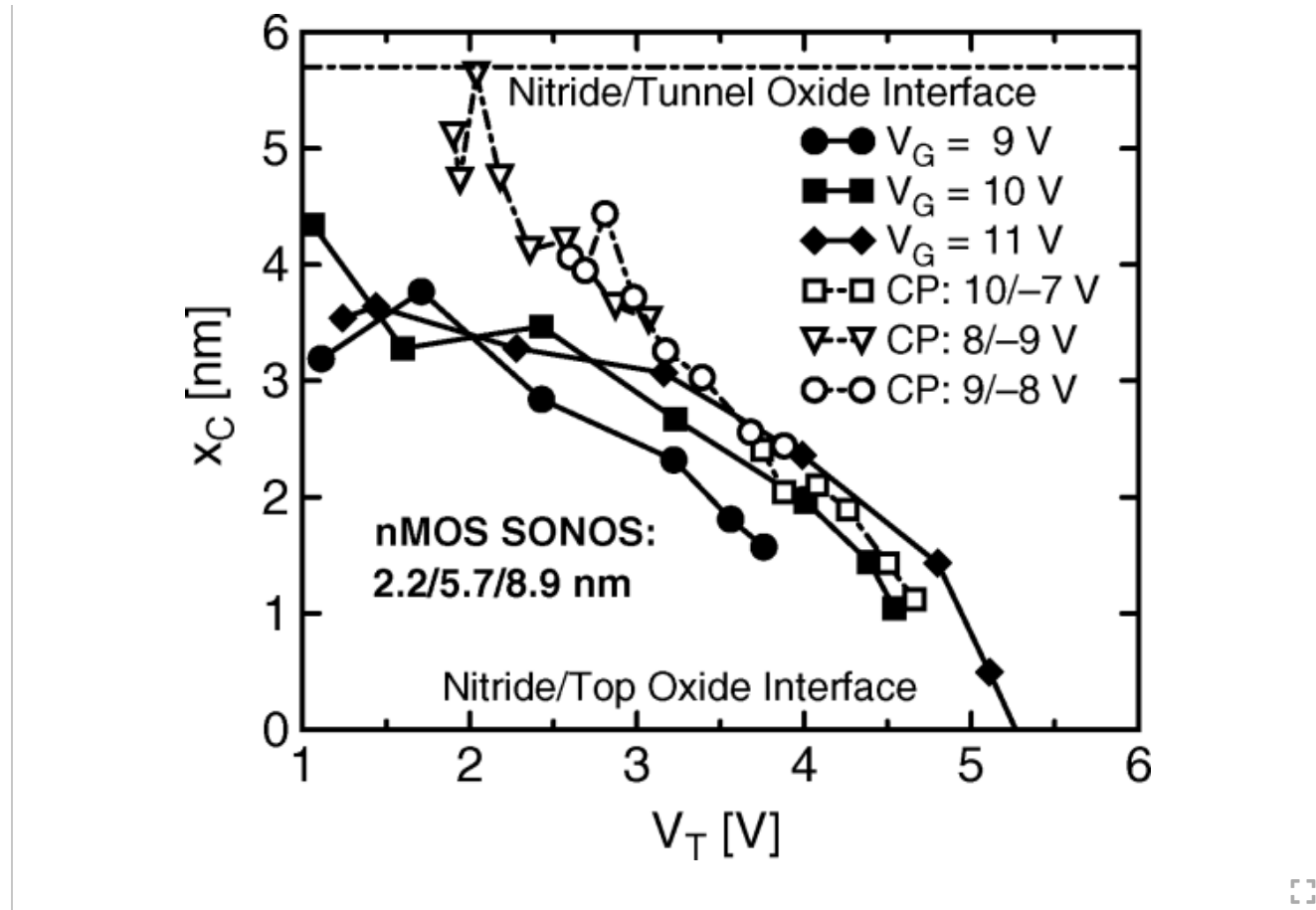


Fig. 8.

Centroid x_C for the nMOS SONOS capacitor (2.2/5.7/8.9-nm ONO stack) extracted during programming. The data obtained with the (filled symbols) direct-measurement technique summarized by (2) are in good agreement with those provided by the (open symbols) LFCP experiments.

Fig. 9 shows the centroid x_C extracted during the programming of SONOS capacitors with different top oxide [$t_{\text{TOP}} = 5.5$ nm; (a)] and silicon nitride thicknesses [$t_N = 3.3$ nm; (b)], whereas all other stack dimensions are the same, as shown in Fig. 7. Also, in this case, we find that x_C is located near the middle of the nitride (slowly shifting toward the top oxide as programming continues) and no gate bias dependence is observed. Moreover,

there is no significant dependence of the measured x_C on the polygate type (open and filled symbols). [Fig. 10](#) shows the x_C values extracted from the SHINOS cells featuring the HfO_2 top dielectric ([Fig. 3](#), device B). Also, in this case, x_C is located inside the nitride, no dependence on V_G is observed, and the data for program and erase are consistent with each other. The results obtained for the pMOS SONOS devices are very similar [\[14\]](#).

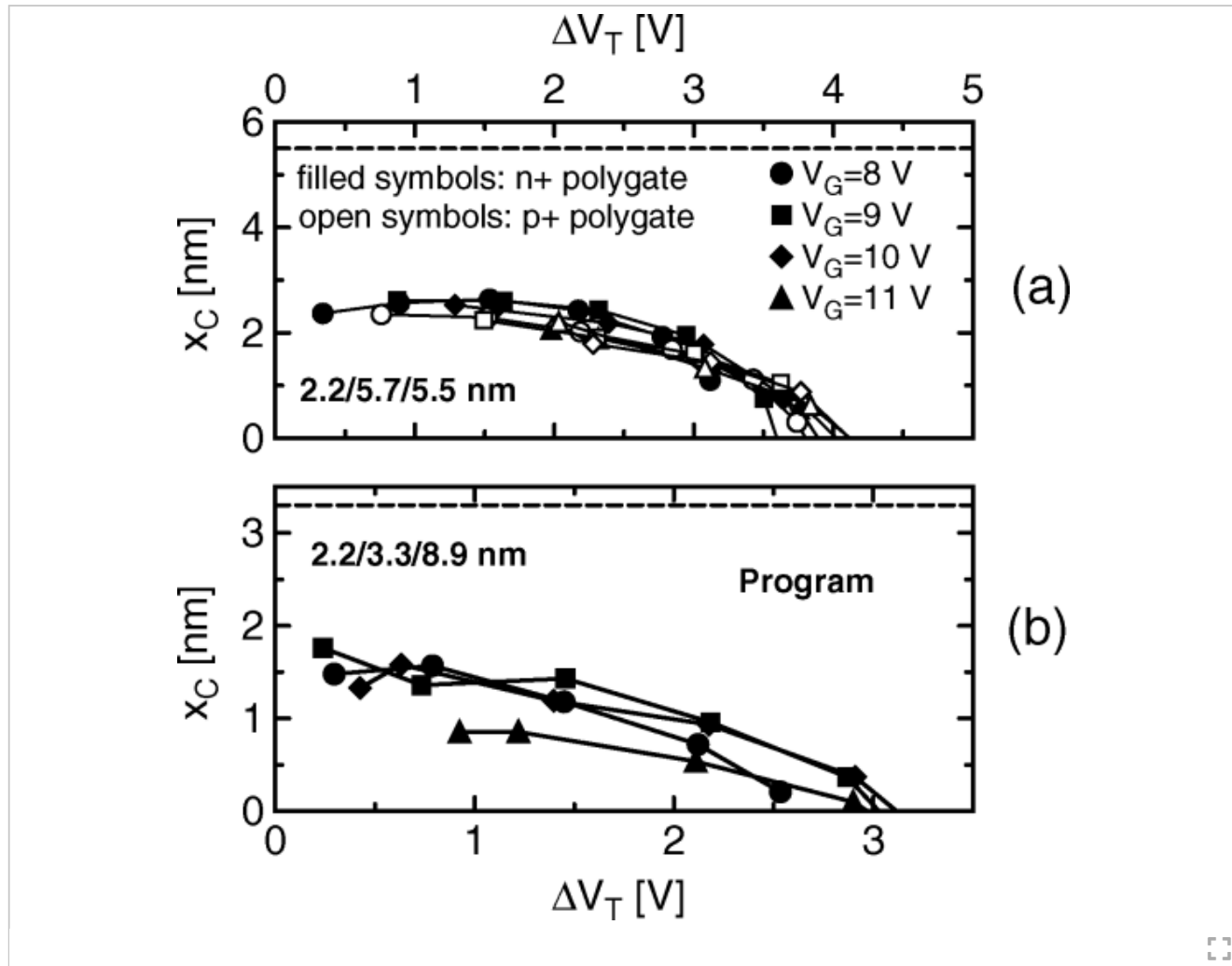


Fig. 9.

Centroid x_C extracted during the programming of SONOS devices, featuring two different ONO stacks: (a) 2.2/5.7/5.5 and (b) 2.2/3.3/8.9. Filled and open symbols are the data for SONOS

with n^+ polygate and p^+ polygate, respectively. The dashed line indicates the tunnel-oxide/silicon-nitride interface.

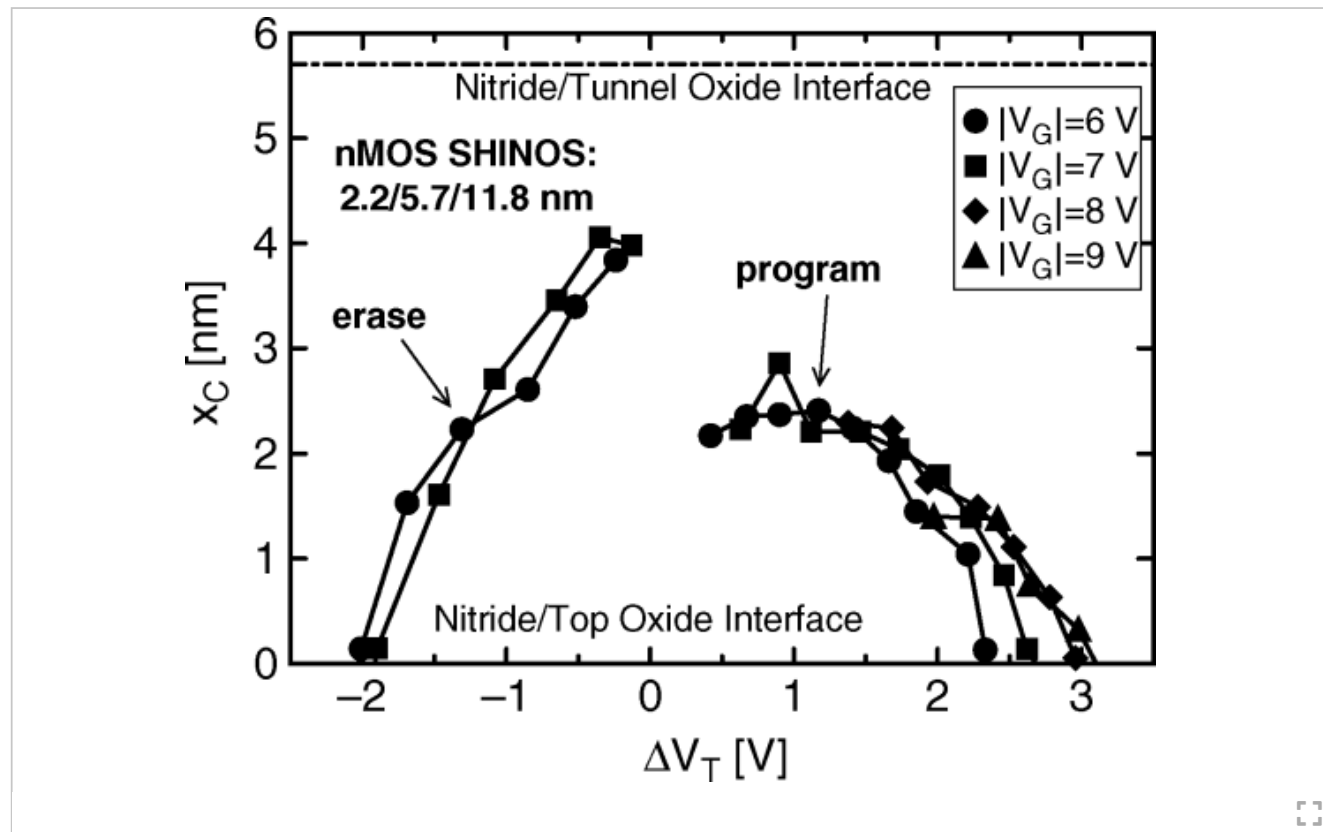


Fig. 10.

Charge centroid extracted during the program and the erase at different V_G 's for the nMOS SHINOS cells, featuring a HfO_2 top oxide of 11.8 nm.

As a summary, [Fig. 11](#) shows the centroid position normalized to the nitride thickness t_N for all the measured SONOS devices. We see how, for moderate ΔV_T values, x_C is always located near the center of the nitride irrespective of the gate-stack composition and the substrate doping type.

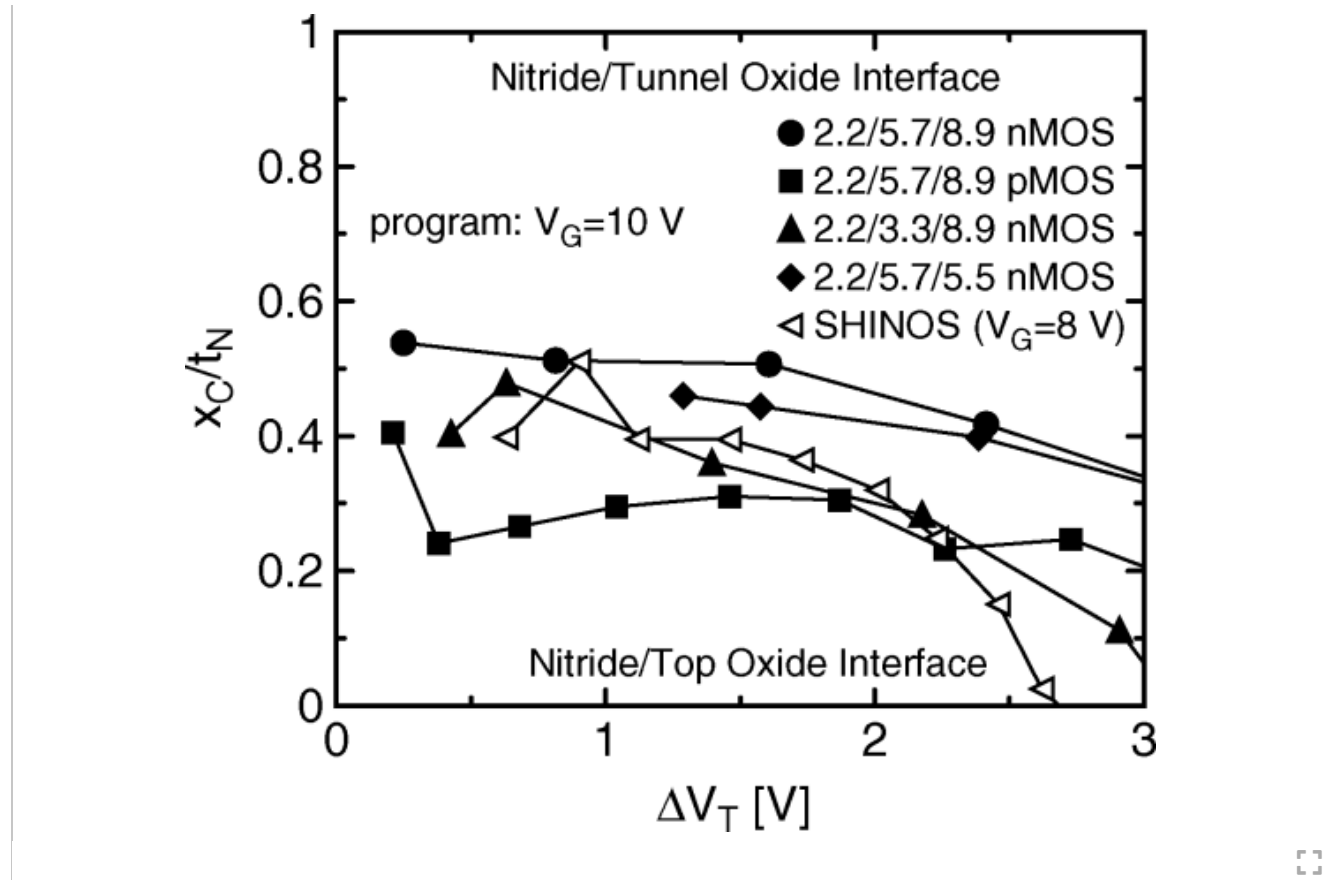


Fig. 11.

Normalized position along the nitride layer of the charge centroid extracted from different devices. Note that the relative position of x_C is near the center of nitride for all the samples, as long as ΔV_T is not very large.

B. Experimental Results for SONONOS Devices

We applied the charge-centroid-extraction procedure also to the SONONOS devices featuring the gate-stack structure shown in Fig. 3 (device C). Fig. 12 shows the P/E characteristics of this SONONOS capacitor, which show a faster programming and erasing compared with SONOS cells (see Fig. 6). In order to perform the x_C extraction accurately, particular care has been taken in the ΔV_T measurement. In fact, the application of even small V_G values produces appreciable threshold-voltage shifts, due to the very thin tunnel oxide (1.3 nm). This results in a considerable hysteresis in the C - V characteristics obtained, spanning a relatively large V_G range

[Fig. 13(a)]. In our capacitors, the ΔV_T detection is performed by measuring the shift of the $C-V$ characteristic, and this can be accomplished by detecting the capacitance at different points, hence V_G values. However, in the SONONOS device, we had to choose an appropriate evaluation point for the ΔV_T [Fig. 13(a), point B] in order to avoid the application of too large ($V_G - V_T$) values that induce a spurious programming during the ΔV_T extraction, as in the case of the choice of the point A in the $C-V$ curve in Fig. 13(a). In the latter case, the gate disturbance leads to a parasitic threshold-voltage shift that depends on the measurement time [Fig. 13(b), in squares]. The experimental results presented in the following have been extracted by detecting the ΔV_T at (triangles) point B, where the gate disturbance is essentially negligible.

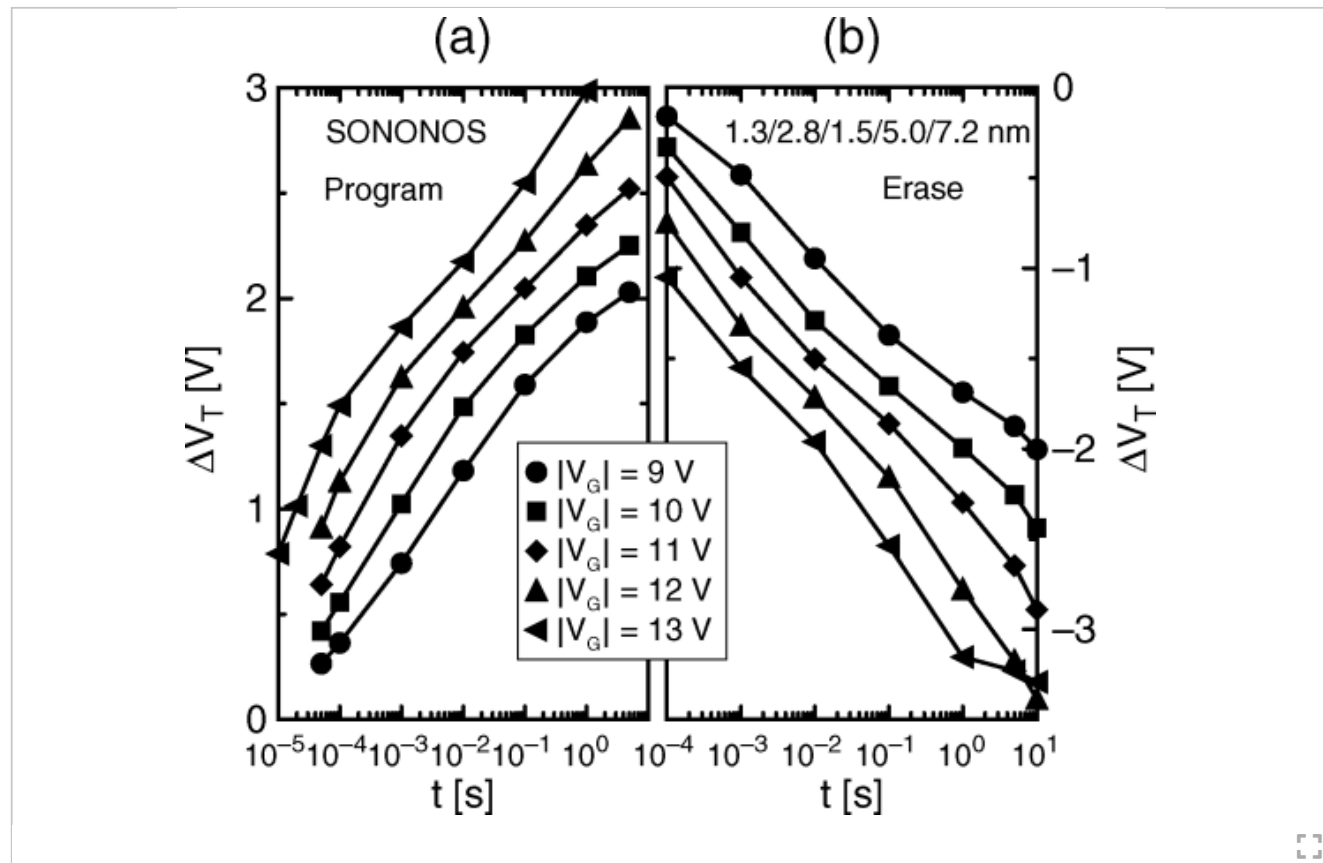


Fig. 12.

(a) Program and (b) erase characteristics of the SONONOS cell for different V_G values.

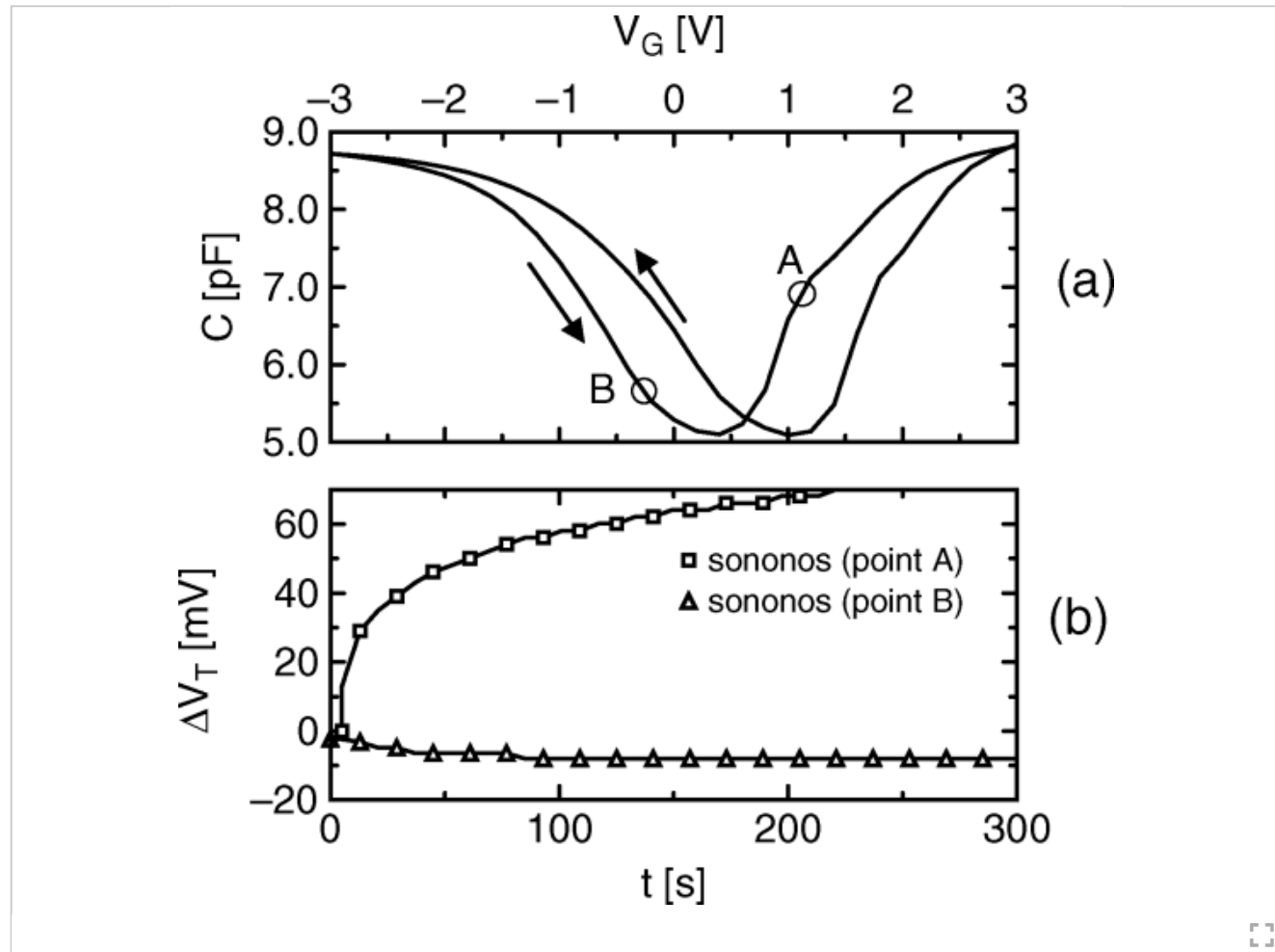


Fig. 13.

(a) C - V curves of the SONONOS, showing a large hysteresis during the measurement. The arrows indicate the direction of the V_G sweep. (b) The ΔV_T value is measured as a shift of the C - V curve. The selection of the right evaluation point in the C - V curve avoids a spurious V_T shift induced by a gate disturb during the detection of the ΔV_T .

On the other hand, it is worth noting that, in the SONOS samples, the discussed issue is not present due to the fact that the C - V curve is not hysteretic because of the thicker tunnel oxide. In this case, the ΔV_T detection can be done at any point of the C - V curve, because the application of the V_G bias that is necessary for the

capacitance measurement does not also produce spurious threshold-voltage shifts if this V_G value is applied to the device for long times.

The x_C values extracted from the SONONOS cells are shown in [Fig. 14](#). Also, in this case, there is no dependence on the applied P/E conditions, and the data for program and erase are consistent. However, unlike standard SONOS devices, SONONOS structures exhibit a centroid x_C that is initially located very close to the second oxide layer ([Fig. 3](#), device C). This is probably due to the fact that the second oxide has been formed by the reoxidation of part of the first nitride layer. This oxide-fabrication method is known to produce a large number of traps at the interface between the first silicon nitride layer and the second oxide layer [\[22\]](#)–[\[24\]](#). The density of these interface traps can be larger than the concentration in the bulk of the nitride layer, and the results in [Fig. 14](#) are consistent with the assumption that the carriers are mainly trapped at the $\text{Si}_3\text{N}_4/\text{SiO}_2$ interface at the beginning of the program/erase operation, namely, for small $|\Delta V_T|$ values. Because in this case, x_C is located in the SiO_2 layer that has a different permittivity ϵ_{OX} with respect to nitride, for the evaluation of the x_C values, [\(2\)](#) has been modified as follows:

$$x_C = -\epsilon_{\text{OX}} \left(\frac{\Delta V_T A}{Q_N} + \frac{t_N}{\epsilon_N} \right) + t_N - t_{\text{TOP}}. \quad (3)$$

[View Source](#) 

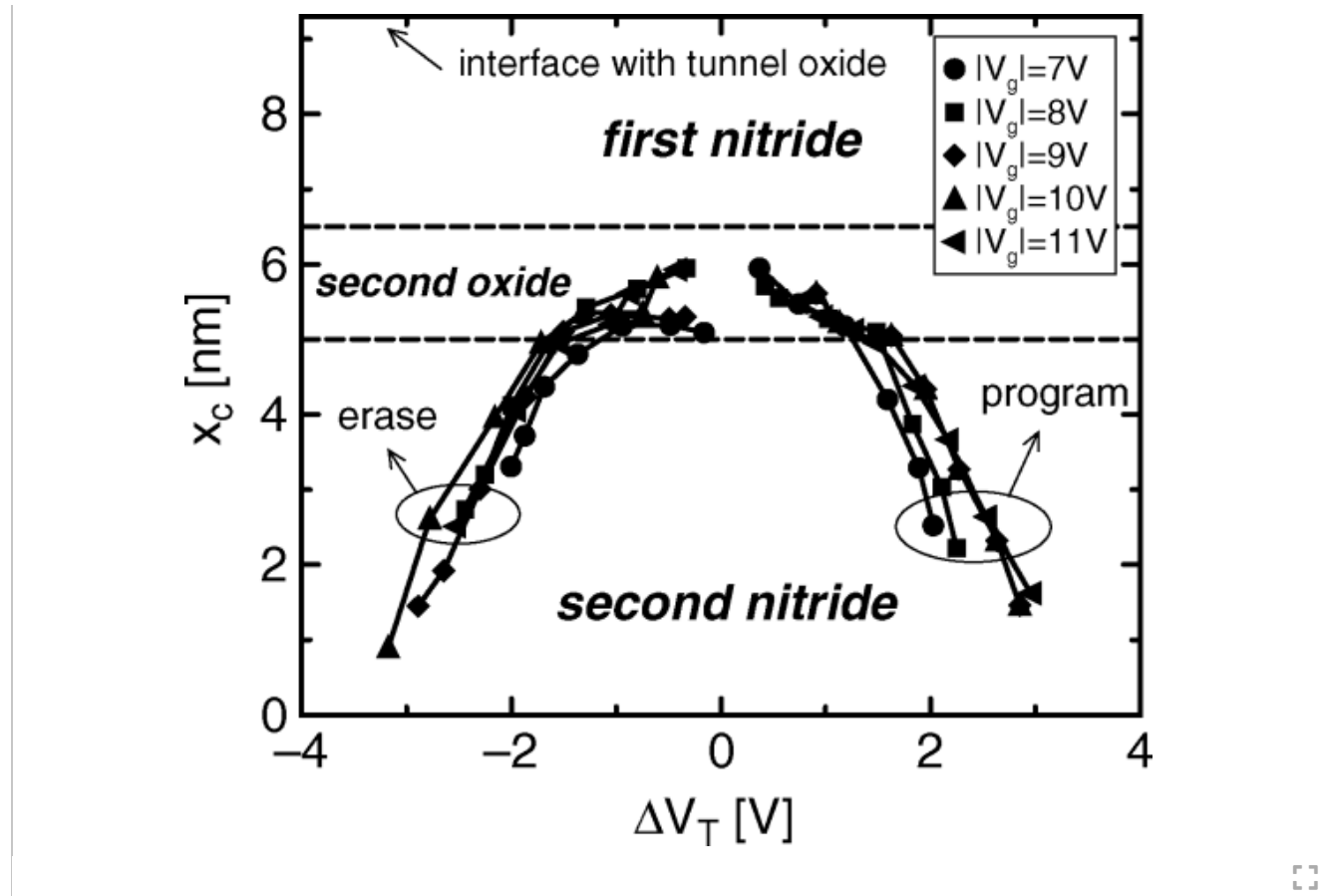


Fig. 14.

Charge centroid extracted from the SONONOS devices. Initially, the charge is trapped near the interface between the first nitride and the second SiO_2 layer, and then, it fills the traps in the center of the second nitride. See [Fig. 3](#) (device C) for the gate-stack definition.

By increasing the program (erase) level, electrons (holes) start filling the bulk traps located in the second nitride film, and x_c moves to the middle of the nitride thickness as in the case of SONOS structures. It seems, however, that no significant amount of charge is trapped in the first nitride layer, although a comparable concentration of free traps should be available with respect to the second nitride layer.

These results also indicate that our experimental technique is indeed sensitive; in fact, it can detect changes of the x_C value due to different trapping properties in the gate stack.

SECTION VI.

Conclusion

We have extensively used carrier-separation experiments with direct measurement of the charge injected during the program and erase operations of SONOS memory cells to study the type of carrier that is mainly involved in each operation as well as the location of the trapped charge Q_N . Preliminary experiments showed that, whereas Q_N is entirely due to electrons during program, during erase, there is a significant contribution of holes, which follows a $(V_G - V_{TP})$ dependence.

Concerning the extraction of the charge centroid, a wide set of experiments in several P/E conditions on SONOS cells with different gate-stack structures has consistently shown that the centroid of the cumulative charge (x_C) is always located close to the center of the trapping layer and it slowly shifts toward the top oxide irrespective of the applied P/E voltages. Moreover, the x_C values do not significantly depend on the top oxide and nitride thicknesses, the composition of the top dielectric, or the substrate and gate doping type.

The centroid results for the SONONOS gate stack instead show that the carriers are initially trapped near the intermediate oxide layer likely because the reoxidation of the first nitride layer has generated a large number of traps at the $\text{Si}_3\text{N}_4/\text{SiO}_2$ interface [22]–[24]. For larger P/E levels, the carriers start filling the bulk traps of the second Si_3N_4 layer, so that the centroid moves toward the middle of the second nitride layer, which is consistent with the results obtained for SONOS devices. Also, for these devices, the behavior of x_C is essentially independent of the P/E conditions.

This large set of experimental data demonstrates the sensitivity of the characterization procedure proposed in this paper, which is thus useful to study the trapping properties of different gate-stack structures and support the design of the nitride-based NVM cells. The data also represent a valid reference for the calibration of program/erase and retention models of Si nitride-based NVM [25].

ACKNOWLEDGMENT

The authors would like to thank Prof. L. Selmi for the fruitful discussions and the constant encouragement. Parts of the samples were fabricated in the framework of the MEDEA+ project NEMeSyS.

Authors	▼
Figures	▼
References	▼
Citations	▼
Keywords	▼
Metrics	▼

[Back to Results](#)

IEEE Personal Account	Purchase Details	Profile Information	Need Help?	Follow
CHANGE USERNAME/PASSWORD	PAYMENT OPTIONS VIEW PURCHASED DOCUMENTS	COMMUNICATIONS PREFERENCES PROFESSION AND EDUCATION TECHNICAL INTERESTS	US & CANADA: +1 800 678 4333 WORLDWIDE: +1 732 981 0060 CONTACT & SUPPORT	f @ in ▶

© Copyright 2025 IEEE - All rights reserved, including rights for text and data mining and training of artificial intelligence and similar technologies.

Ex. D Abstract

Bit Cost Scalable Technology with Punch and Plug Process for Ultra High Density Flash Memory

Publisher: IEEE

Cite This

 PDF

H. Tanaka ; M. Kido ; K. Yahashi ; M. Oomura ; R. Katsumata ; M. Kito

All Authors

457
Cites in
Papers

538
Cites in
Patents

10974
Full
Text Views



Abstract

Document Sections

» Introduction

» Concept

» Fabrication

» Result and Discussion

» Conclusion

Authors

Figures

References

 Citations

Abstract:
We propose Bit-Cost Scalable (BiCS) technology which realizes a multi-stacked memory array with a few constant critical lithography steps regardless of number of stacked layer to keep a continuous reduction of bit cost. In this technology, whole stack of electrode plate is punched through and plugged by another electrode material. SONOS type flash technology is successfully applied to achieve BiCS flash memory. Its cell array concept, fabrication process and characteristics of key features are presented.

Published in: 2007 IEEE Symposium on VLSI Technology

Date of Conference: 12-14 June 2007

DOI: 10.1109/VLSIT.2007.4339708

Date Added to IEEE Xplore: 08 October 2007

Publisher: IEEE

Print ISBN:978-4-900784-03-1

Conference Location: Kyoto, Japan

▼ **ISSN Information:**

Introduction

Keywords

Metrics

More Like This

It is strongly demanded to keep a trend of increasing bit density and reducing bit cost of flash memories. To achieve that, aggressive scaling of the device dimension and multi-level-cell have been developed. However, bit cost will turn to rise up near future, as shown in [figure 1](#), because process cost will increase more rapidly than shrink rate.

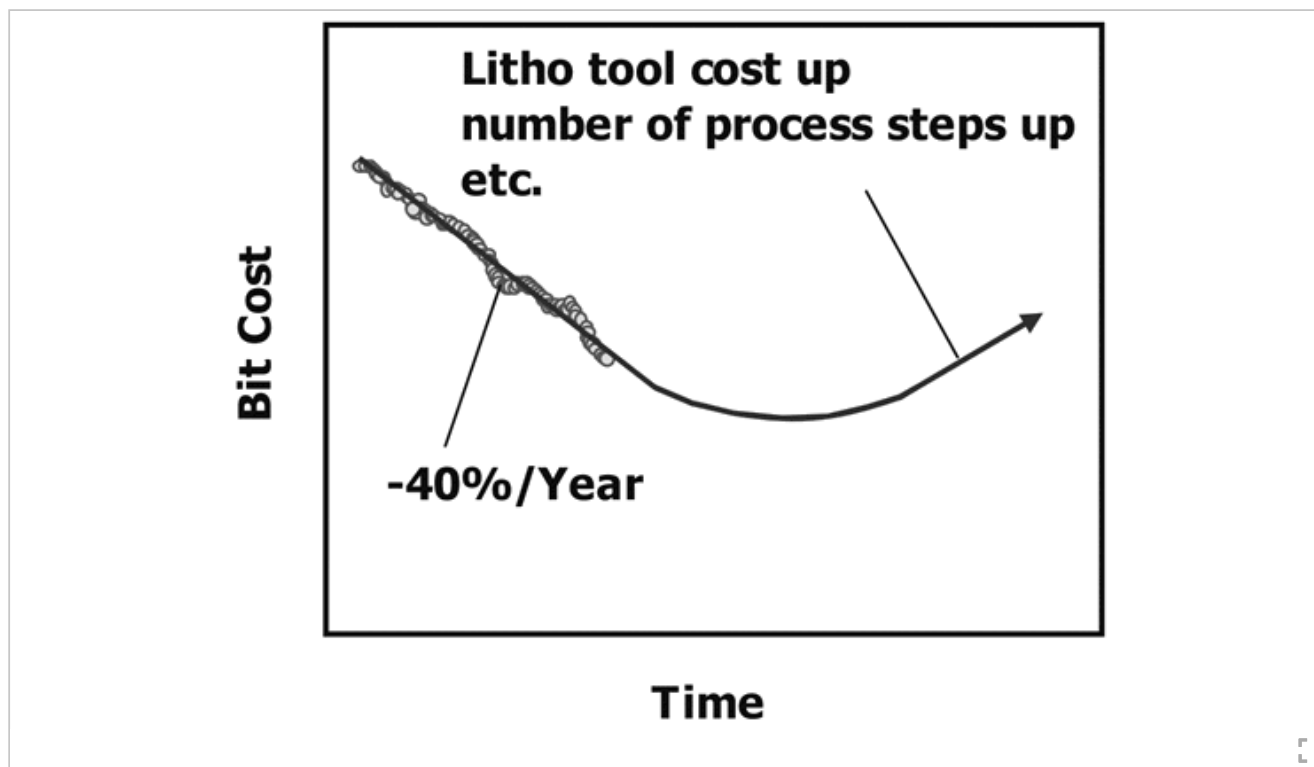


Fig. 1 Bit cost trend of flash memory.

One solution to avoid such situation is three dimensionally stacked array structures as reported in previous papers [1] [2], in which planar NAND structures are simply stacked. But there are two severe issues on those structures. One is that several critical lithography steps to make minimum feature size are necessary for each stacked layer. The other is that the area of driver transistors for control gates is multiplied by the number of layers. Dotted line in the [figure 2](#) illustrates the estimated bit cost of three dimensionally stacked NAND, calculated by a simple formula inset in [figure 2](#). It is important to note that the cost does not decrease or even increases due to a yield loss and an area penalty if the number of stacked layers is more than three.

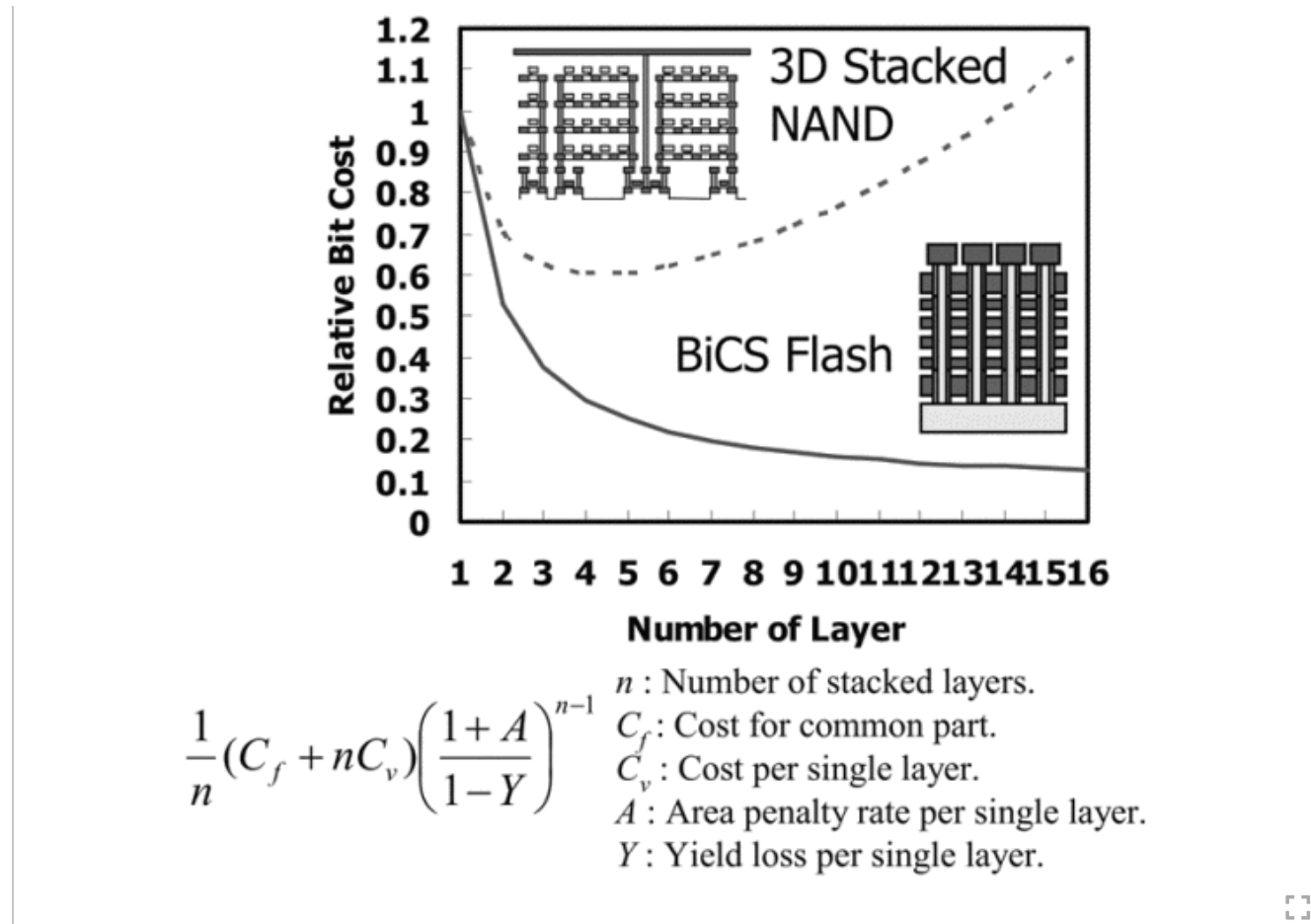


Fig. 2 Bit Cost scalability of three dimensional flash memory.

In this paper, a novel SONOS flash memory using a newly devised Bit-Cost Scalable (BiCS) technology which realizes a multi-stacked array with a few constant number of critical lithography process steps, without area penalty and with continuous reduction of bit cost as shown as solid line in [figure 2](#) is proposed.

Concept

Figures 3(a), 3(b) and figure 4 show schematic Birds-eye view, top down view of BiCS flash memory array and the equivalent circuit, respectively. String of the cells is on the plugs located vertically in the holes punched through whole stack of the gate plates. Each plate acts as control gate except the lowest plate which takes a role of the lower select gate. Cell size of BiCS memory is $6F^2/n$, where n is the number of control gate plates. Bit density can be increased by adding more gate plates, while the number of the critical lithography steps remains constant because whole stack of control gates is punched with only one lithography step. A single cell is accessed by control gate on the string which is selected by a bit line and an row select line. The bottom of memory plug is connected to source diffusion formed on the silicon substrate. For the erase operation of BiCS Flash memory, hole current which is generated by GIDL near the lower select gate is used. Edges of control gate are processed to be stair-like structure for via holes which are connected to driver transistors. They are shared by several rows of strings, so that the area for driver transistors does not increase even if more control gates are stacked. Table I summarizes the comparison of BiCS Flash and simply stacked 3D NAND.

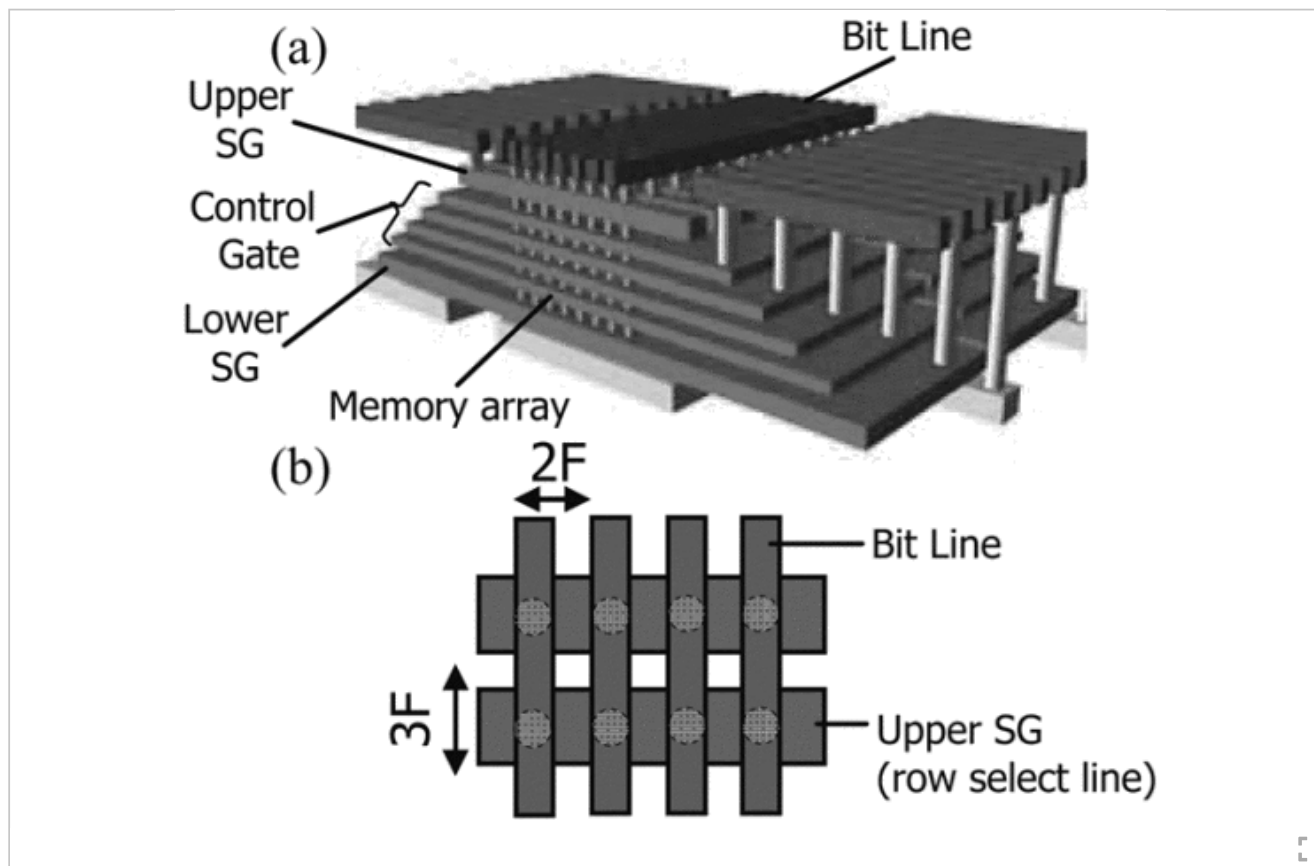


Fig. 3 (a) Birds-eye view of BiCS flash memory, (b) Top down view of BiCS flash memory array.

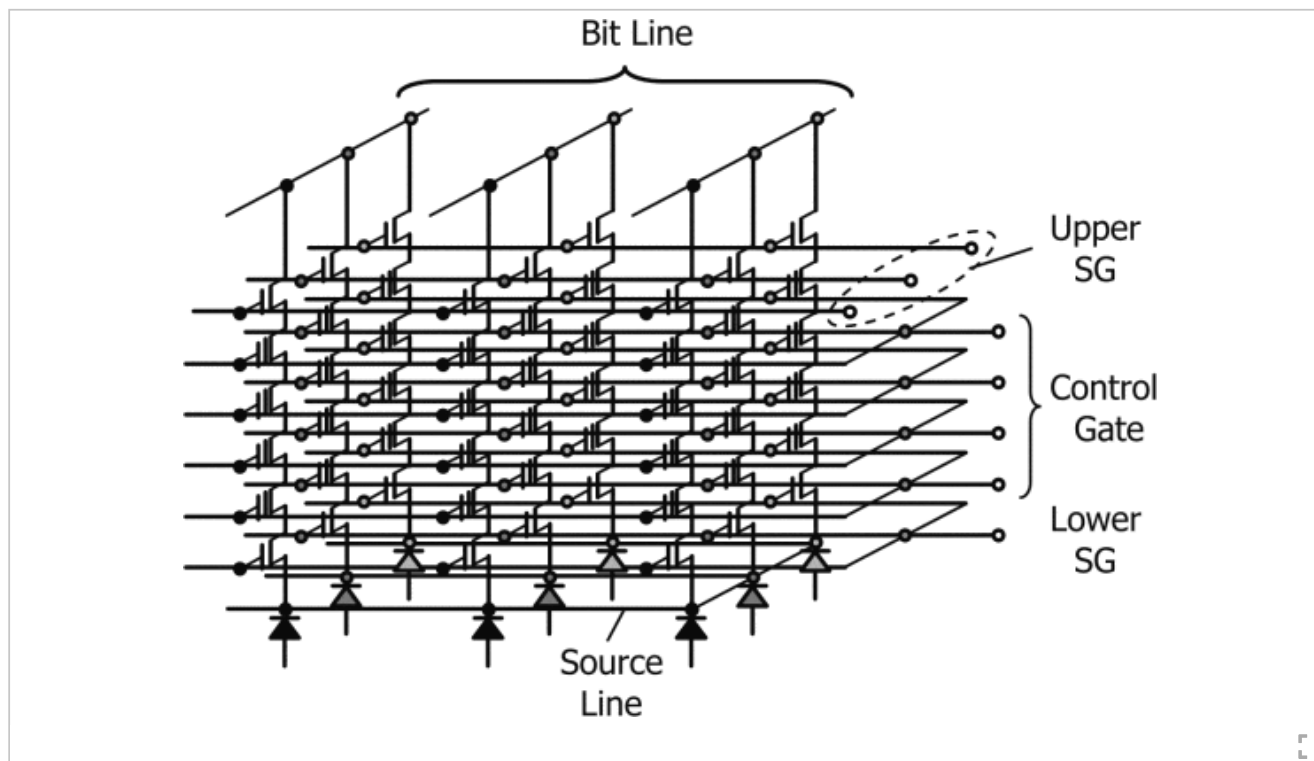


Fig. 4 Equivalent circuit of BiCS flash memory.

TABLE I Comparison of three dimensional flash memories.

	Pros	Cons
3D Stacked NAND	Expansion of Current Technology	Higher Cost (Large number of Lithography Steps)
BiCS Flash	Lower Cost (Small number of Lithography Steps)	New Technology - LPCVD tunnel oxide - Vertical FET - Multi Gate Stack RIE



Fabrication

Figure 5 shows the fabrication sequence. Lower select gate transistor, memory string and upper select gate transistor are fabricated individually. Gate material is P+ poly-Si. Holes for transistor channel or memory plug are punched through and LPCVD TEOS film or ONO films are deposited. The bottom of dielectric films are removed by RIE and plugged by amorphous Si. Arsenic is implanted and activated for drain and also source of upper device. Figure 6 shows schematic cross sections of vertical FET and SONOS memory cell on poly-Si gate. It should be pointed out that ONO films are deposited in the opposite order of the conventional SONOS device, i.e. from LPCVD TEOS film as top block oxide (5nm), LPCVD SiN film (11nm) and LPCVD TEOS film as tunnel oxide (2.5nm). Edges of control gate are processed into stair-like structure by repeating of RIE and resist sliming as shown in figure 8. For minimizing disturb, whole stack of control gate and lower select line are etched to have a slit which separates a block of memory plugs from each other. Only upper select gate is cut into line pattern to work as row address selector. Via hole and BL are processed on the array and peripheral circuit simultaneously.

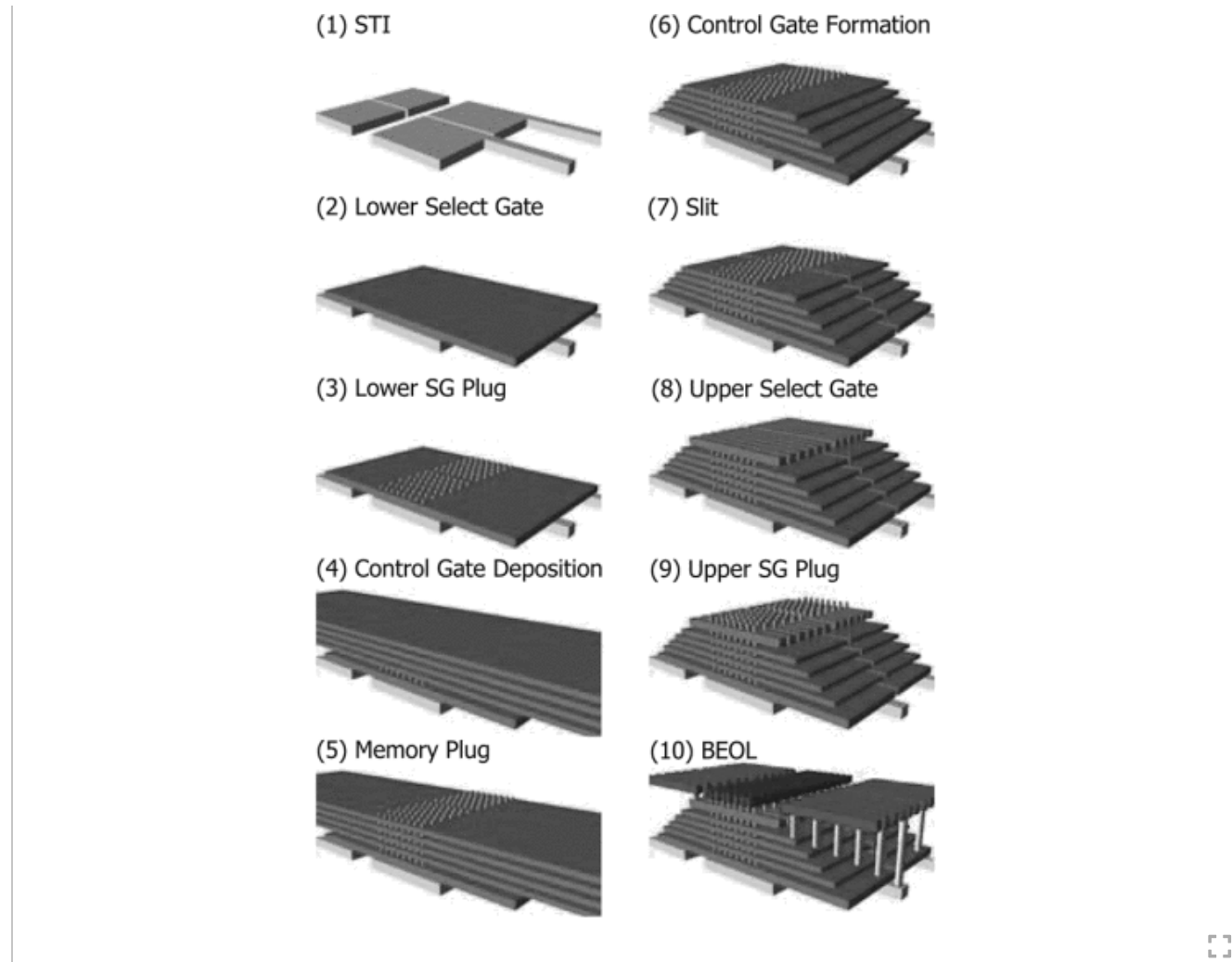


Fig. 5 Fabrication sequence of BiCS flash memory.

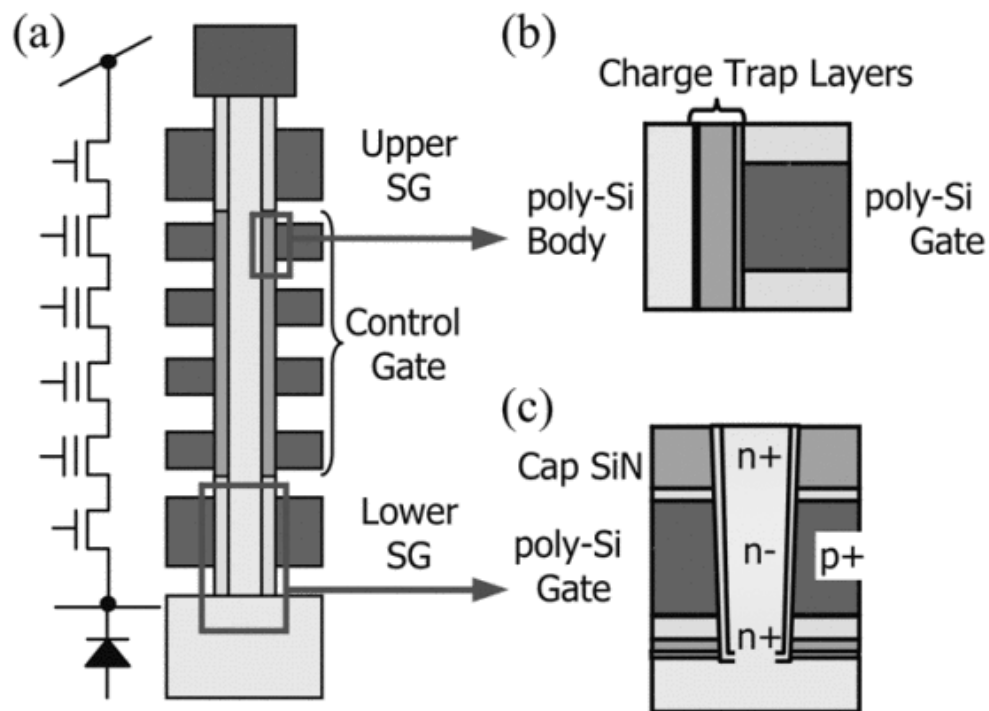


Fig. 6 (a) Cross section of BiCS flash memory string, (b) Cross section of vertical SONOS cell, (c) Cross sections of vertical FET.

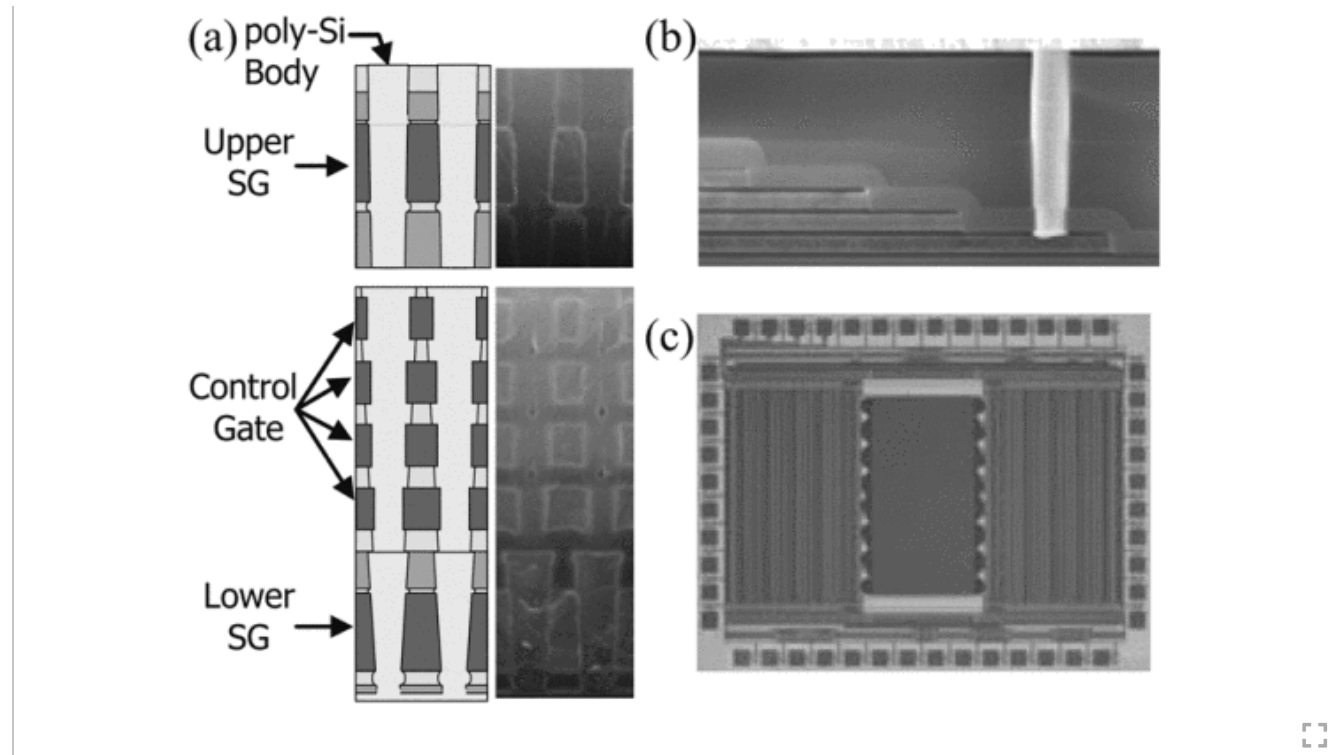


Fig. 7 (a) Cross sectional SEM of BiCS flash memory string, (b) Cross sectional SEM of edge of control gates, (c) $n \times 512$ kbit macro image.

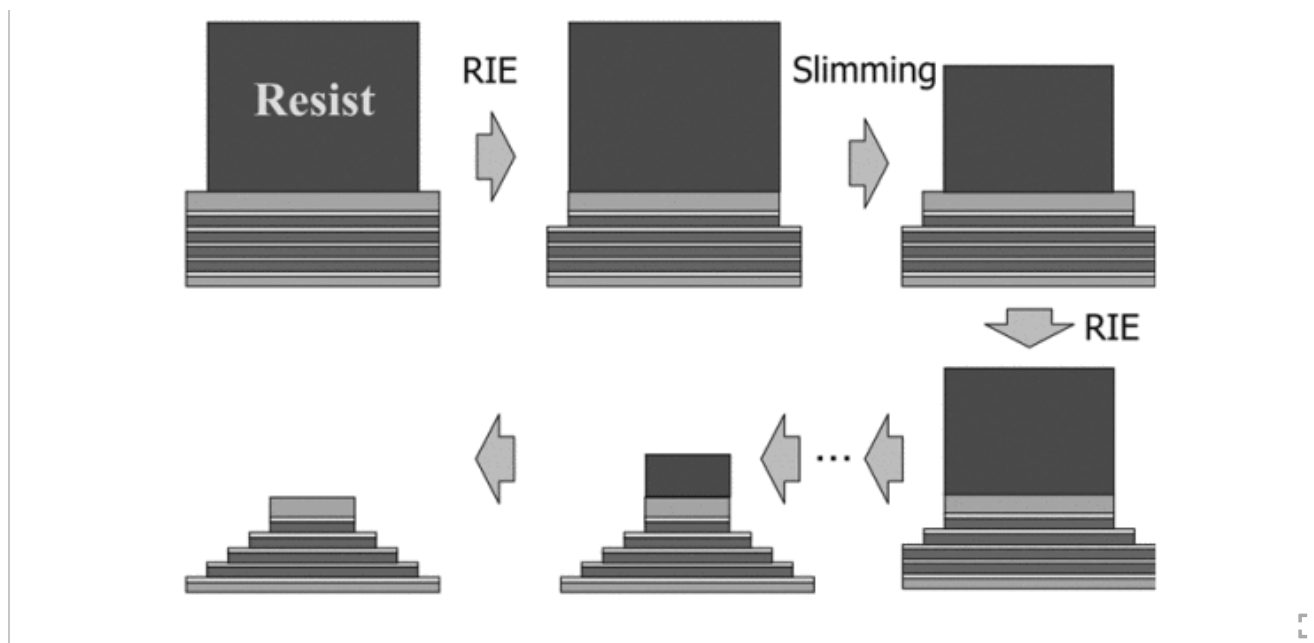


Fig. 8 Fabrication Sequence of edge of control gates into stair-like structure.

Result and Discussion

Figures 7(a) and 7(b) show the cross sectional SEM images of the successfully fabricated 90nm BiCS flash cell array and the stair-like structure at the edge of control gates. Figure 7(c) is a top down photograph of $n \times 512k$ bit BiCS Flash macro.

I_d - V_g characteristics of select gate FET ($L = 150\text{nm}$, $\Phi = 90\text{nm}$) is shown in figure 9. Subthreshold slope is 250 mV/dec. In the case that I_{off} is 30 pA, on-current is $2.4\mu\text{A}$. Figure 10 shows I_d - V_g characteristics of the vertical SONOS FET after program and erase. Figure 11 shows the program/erase characteristics of SONOS FET. Endurance to $1E5$ cycles is shown in figure 12. V_{th} shift is less than 0.5V after $1E4$ cycle. Data retention measured on SONOS FET at initial and after 1000 cycles of endurance are shown in figure 13. 2.5V of V_{th} window is remains after 10years.

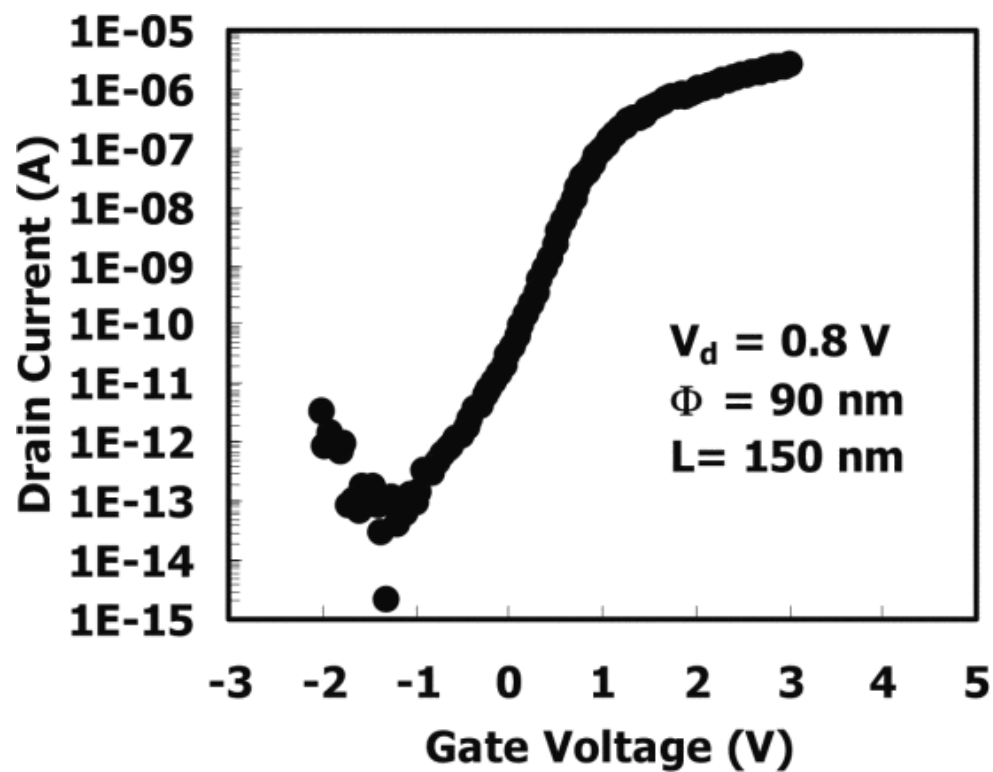


Fig. 9 I_d - V_g characteristics of vertical transistor.

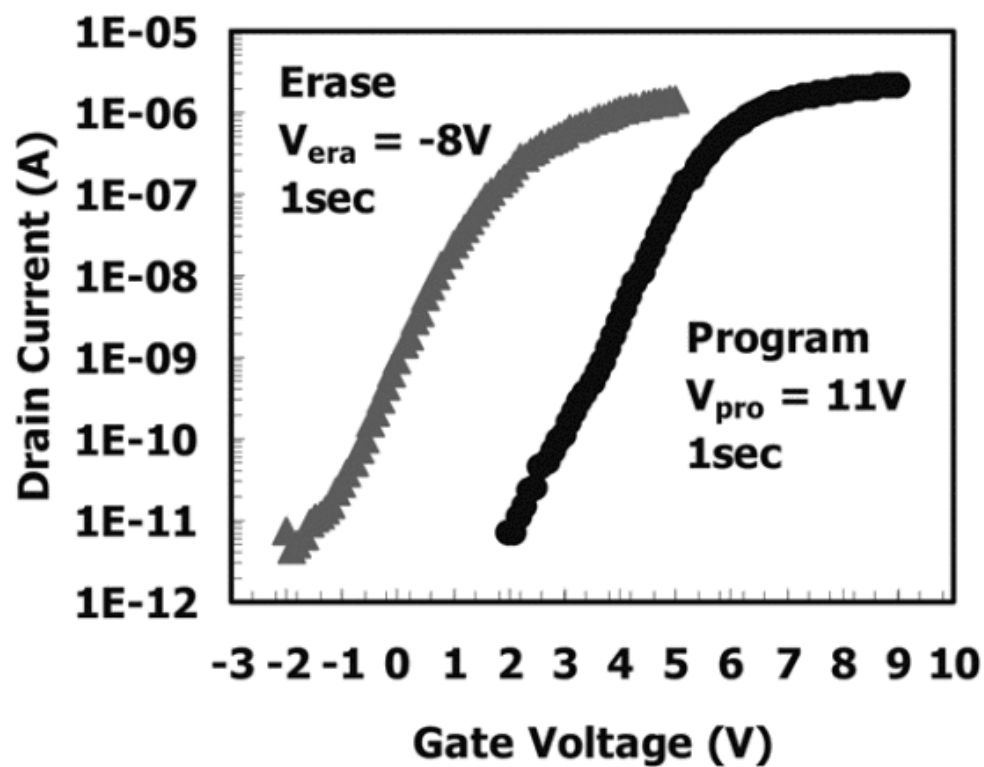


Fig. 10 Id-Vg characteristics for Program/Erase of vertical SONOS.

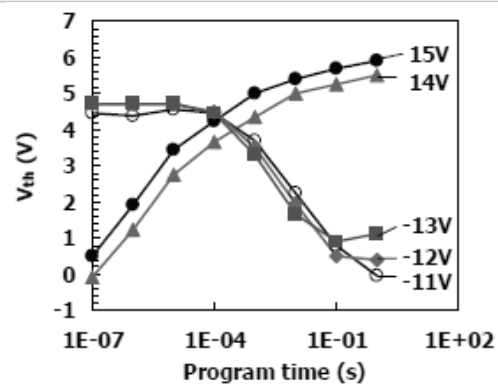
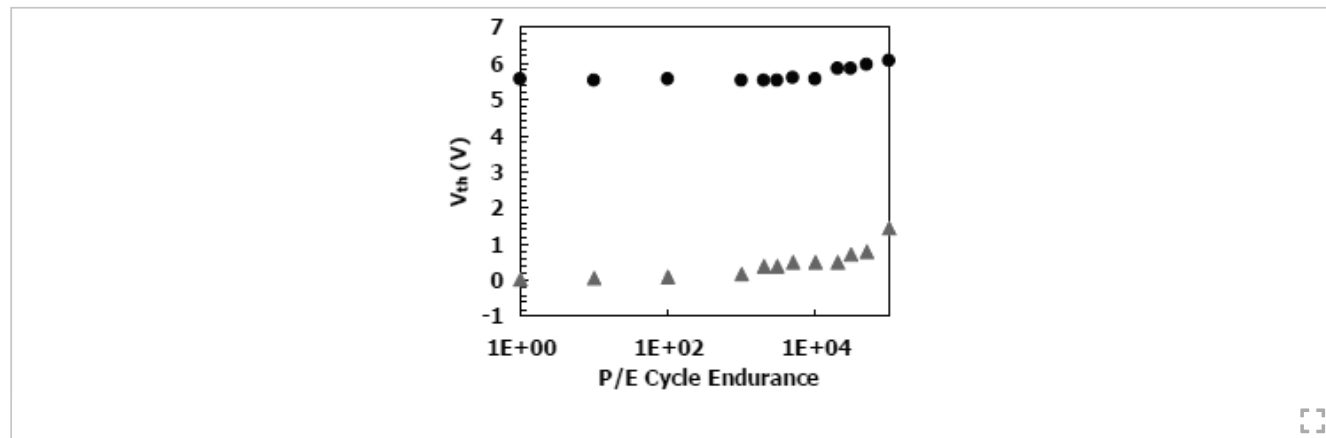


Fig. 11 Program/Erase characteristics.**Fig. 12** Endurance of vertical SONOS.

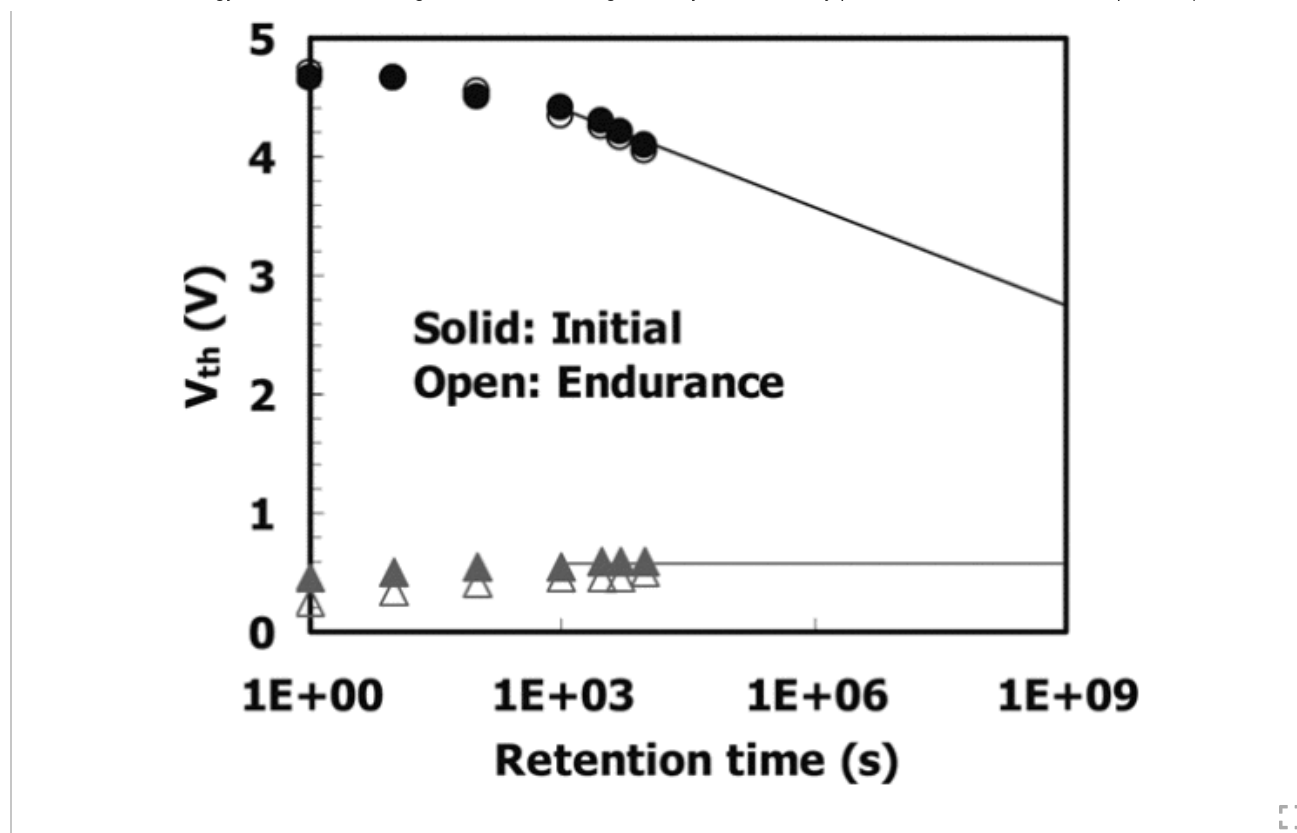


Fig. 13 Data retention of vertical SONOS.

Conclusion

We have proposed BiCS flash memory as one of the most cost-effective memories to increase bit density. We have confirmed that this technology is a promising candidate for the Ultra High Density Memory.

Figures	▼
References	▼
Citations	▼
Keywords	▼
Metrics	▼

[Back to Results](#)

IEEE Personal Account	Purchase Details	Profile Information	Need Help?	Follow
CHANGE USERNAME/PASSWORD	PAYMENT OPTIONS VIEW PURCHASED DOCUMENTS	COMMUNICATIONS PREFERENCES PROFESSION AND EDUCATION TECHNICAL INTERESTS	US & CANADA: +1 800 678 4333 WORLDWIDE: +1 732 981 0060 CONTACT & SUPPORT	f @ in ▶

[About IEEE Xplore](#) | [Contact Us](#) | [Help](#) | [Accessibility](#) | [Terms of Use](#) | [Nondiscrimination Policy](#) | [IEEE Ethics Reporting](#)  | [Sitemap](#) | [IEEE Privacy Policy](#)

A public charity, IEEE is the world's largest technical professional organization dedicated to advancing technology for the benefit of humanity.

© Copyright 2025 IEEE - All rights reserved, including rights for text and data mining and training of artificial intelligence and similar technologies.