

2006 International Electron Devices Meeting TECHNICAL DIGEST

Papers have been printed without editing as received from the authors.

All opinions expressed in the digest are those of the authors and are not binding on The Institute of Electrical and Electronics Engineers, Inc.

Publication of a paper in this Digest is in no way intended to preclude publication of a fuller account of the paper elsewhere.

Copies of available volumes of this Digest may be obtained from The Institute of Electrical and Electronics Engineers, Inc., 445 Hoes Lane, Piscataway, NJ 08855.

Copyright and Reprint Permission: Abstracting is permitted with credit to the source. Libraries are permitted to photocopy beyond the limit of U.S. copyright law for private use of patrons those articles in this volume that carry a code at the bottom of the first page, provided the per copy fee indicated in the code is paid through Copyright Clearance Center, 222 Rosewood Drive, Danvers, MA 01923. For other copying, reprint or republication permission, write to IEEE Copyrights Manager, IEEE Operations Center, 445 Hoes Lane, P.O. Box 1331, Piscataway, NJ 08855-1331. All rights reserved. Copyright ©2006 by the Institute of Electrical and Electronics Engineers, Inc.

IEEE Catalog Number: 06CH37807C

ISBN: 1-4244-0439-8

Library of Congress Number: 81-642284

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 $^{\circ}\text{C}$) 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 μ m
Channel width	0.09 μ m

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

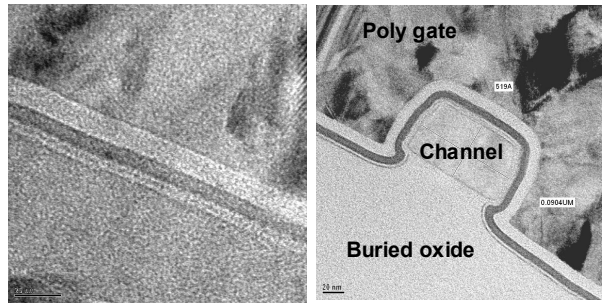


Fig. 1 TEM of the O1/N1/O2/N2/O3 structure for TFT BE-SONOS.

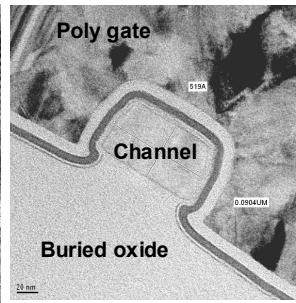


Fig. 2 Tri-gate structure of the TFT device.

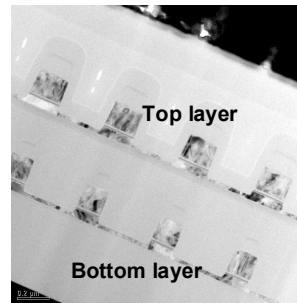


Fig. 3 Channel-length direction of double-layer TFT NAND devices.

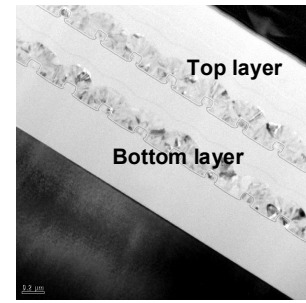


Fig. 4 Channel-width direction of double-layer TFT NAND devices.

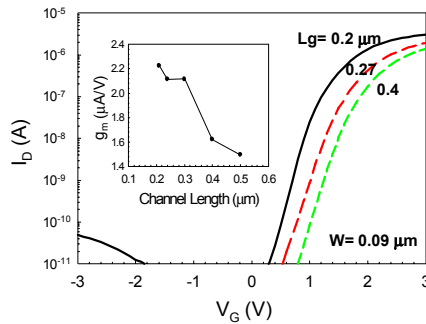


Fig. 5 I_V characteristics for various channel length devices. Inset shows the corresponding transconductance (g_m). The channel width is 0.09 μ m.

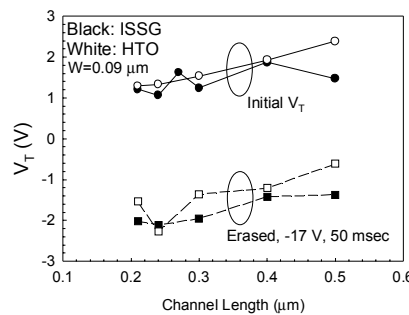


Fig. 6 Short-channel V_T roll-off characteristics for both the initial state and the -FN erased state, with ISSG and HTO top oxide.

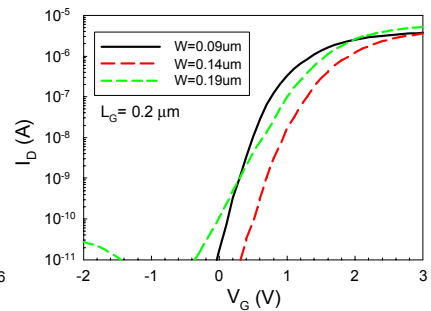


Fig. 7 I_V characteristics for various channel width devices. Wider devices show worse subthreshold behavior. The channel length is 0.2 μ m.

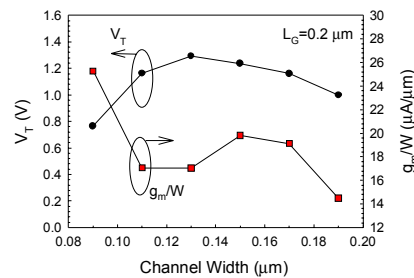


Fig. 8 Narrow-width characteristics. Both normalized transconductance (g_m/W) and V_T are shown. The narrowest device ($W=0.09 \mu$ m) has the best performance.

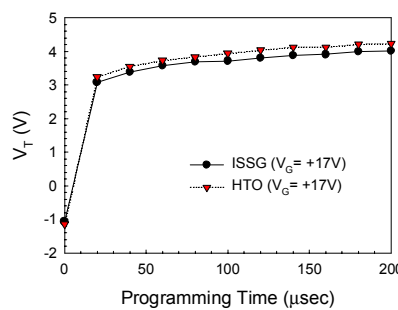


Fig. 9 Comparison of the +FN program characteristics for ISSG and HTO top oxide.

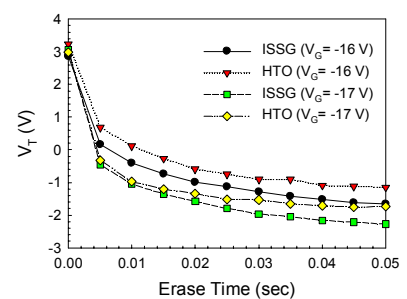


Fig. 10 Comparison of the -FN erase characteristics for ISSG and HTO top oxide.

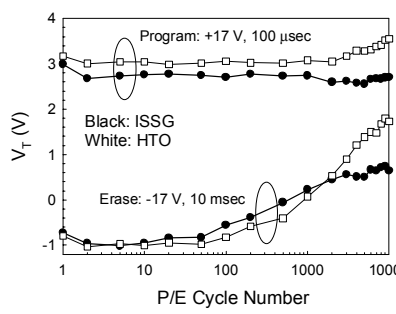


Fig. 11 Dumb-mode P/E cycling endurance (without adjusting the P/E conditions during endurance test). Both ISSG and HTO top oxide are compared.

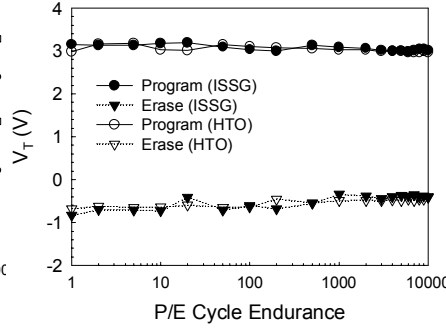


Fig. 12 P/E cycling endurance (with adjusted P/E conditions during endurance test). Both ISSG and HTO top oxide are compared.

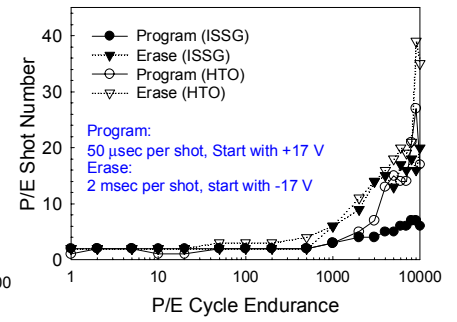


Fig. 13 The required P/E shot number during adjusted P/E tests in Fig. 12. Both ISSG and HTO top oxide are compared.

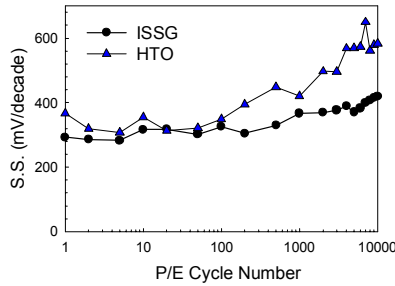


Fig. 14 Subthreshold slope degradation during 10K P/E cycling tests. Both ISSG and HTO top oxide are compared.

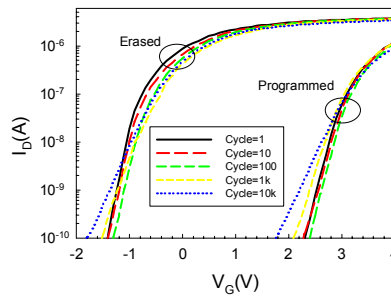


Fig. 15 I_V characteristics of ISSG top oxide device during 10K P/E cycling endurance testing.

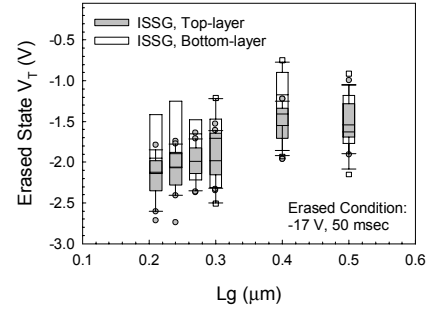


Fig. 16 Short-channel effect comparison of the top and bottom layer TFT NAND devices. A total of 28 identical devices are measured. The devices are in erase saturation state.

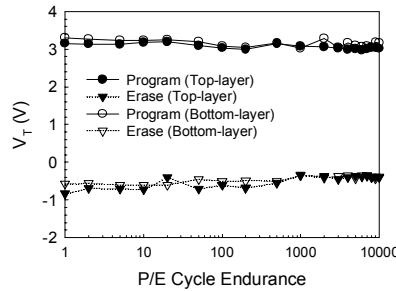


Fig. 17 Adjusted P/E cycling endurance (with adjusted P/E conditions during endurance test). Both bottom and top layers are compared. ISSG top oxide is used.

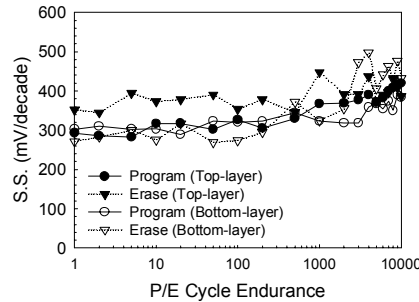


Fig. 18 Subthreshold slope degradation during 10K P/E cycling tests. Both bottom and top layers are compared. ISSG top oxide is used.

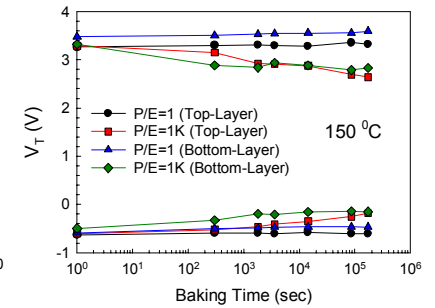


Fig. 19 Retention characteristics. Programmed and erased state with various cycle endurance are compared.

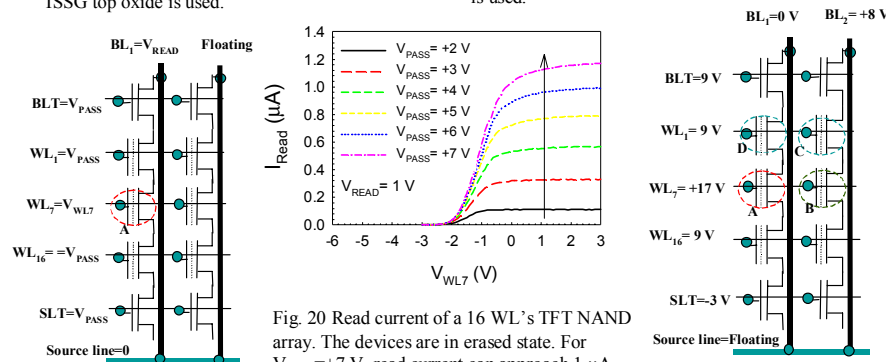


Fig. 20 Read current of a 16 WL's TFT NAND array. The devices are in erased state. For $V_{PASS}=+7$ V, read current can approach 1 μ A, sufficient for NAND Flash design.

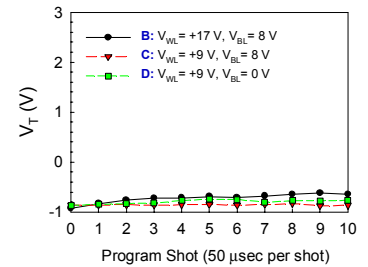


Fig. 21 Program disturb of a TFT NAND array. Cells C and D are under medium field gate disturb, while cell B can be program-inhibited by raising the channel potential.