Tunnel Barrier Engineering for Non-Volatile Memory

Jongwan Jung* and Won-Ju Cho**

***Abstract*—Tunnel oxide of non-volatile memory (NVM) devices would be very difficult to downscale if ten-year data retention were still needed. This requirement limits further improvement of device performance in terms of programming speed and operating voltages. Consequently, for low-power applications with Fowler-Nordheim programming such as NAND, program and erase voltages are essentially sustained at unacceptably high levels. A promising solution for tunnel oxide scaling is tunnel barrier engineering (TBE), which uses multiple dielectric stacks to enhance field-sensitivity. This allows for shorter writing/erasing times and/or lower operating voltages than single SiO_2 tunnel oxide without altering the ten-year data retention constraint. In this paper, two approaches for tunnel barrier engineering are compared: the crested barrier and variable oxide thickness. Key results of TBE and its applications for NVM are also addressed.**

***Index terms*—Non-volatile memory, tunnel barrier engineering, tunnel oxide, retention, VARIOT**

I. Introduction

When tunnel oxide for non-volatile memory (NVM) is scaled, the direct tunneling effect becomes dominant. A simple calculation shows that the minimum thickness is around 6 nm [1]. In addition, the strain-induced leakage current increases for thinner oxides and aggravates scaling of tunnel oxide thickness, as shown in Fig. 1 [2]. In current flash NVM devices, the tunnel oxide thickness is approximately in the 7–8 nm range. This requirement limits further improvement of device performance in terms of programming speed and operating voltages. When the barrier is thin, the program and erase process is rapider but charge leakage destroys the retention time. When the barrier is relatively thick, long charge retention times are achieved but a higher voltage and a longer period of time are required to program and erase the floating gate. Because of this, downscaling of tunnel oxide would be very difficult if ten-year data retention were still required [3]. Consequently, for low-power applications with Fowler-Nordheim (FN) programming, such as NAND, program and erase voltages are essentially sustained at unacceptably high levels. A promising solution for tunnel oxide scaling is tunnel barrier engineering (TBE), which uses multiple dielectric stacks to enhance field-sensitivity [4,5]. The main potential of enhancing field-sensitivity through TBE could allow for shorter writing/erasing times and/or smaller operating voltages than those obtained using single SiO_2 tunnel oxide, and

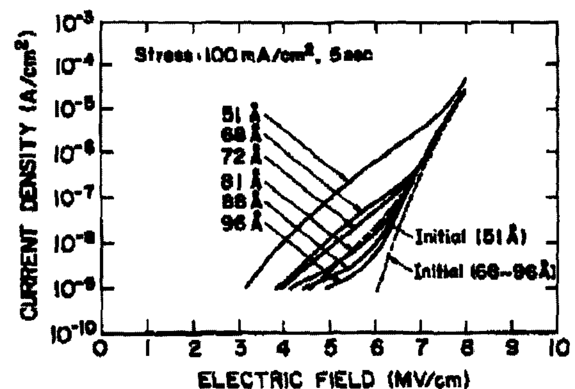


Fig. 1. Current density-electric field characteristics measured by capacitors having 5.1 to 9.6-nm oxide thickness before and after charge injection stress [2].

Manuscript received Feb. 27, 2008; revised Mar. 10, 2008.
* Department of NanoScience and Technology, Sejong University, #98, Gunja-dong, Gwangjin-gu, Seoul 143-747, Korea,
** Department of Electronic Materials Engineering, Kwangwoon University 447-1, Wolgye-dong, Nowon-gu, Seoul, Korea, 139-701,
E-mail : chowj@kw.ac.kr

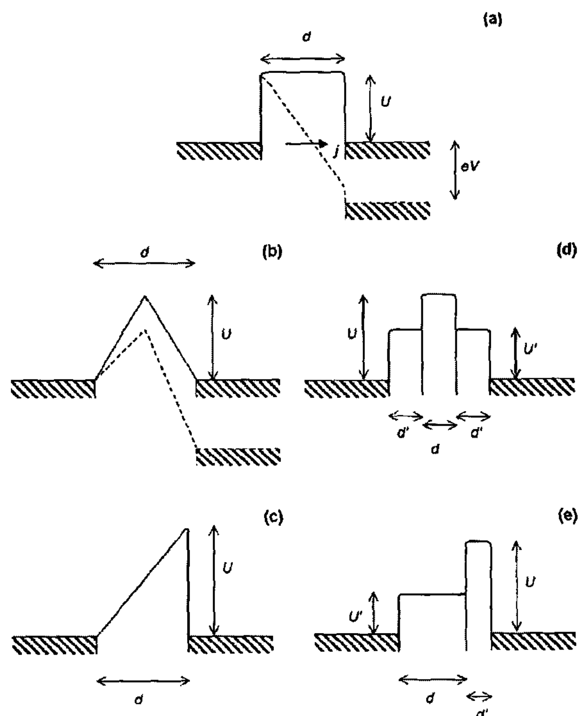


Fig. 2. Conduction band edge diagrams of various tunnel barriers: (a) typical uniform barrier; (b) idealized crested symmetric barrier; (c) idealized asymmetric barrier; (d) crested, symmetric layered barrier; and (e) asymmetric layered barrier. Dashed lines in panels (a) and (b) show barrier tilting caused by applied voltage [4].

the ten-year data retention constraint would not be harmed. Multi-layer engineered tunnel barriers for NVM applications were first introduced by Likharev in 1998 with the concept of crested barriers [4]. The international technology roadmap for semiconductors included engineered tunnel barrier memory in emerging research devices in 2003 [6]. The engineered tunnel barrier includes crested barrier floating gate memory [4] and variable oxide thickness (VARIOT) floating gate memory [5]. Application of TBE was mostly aimed at floating gate memory in its early stages. Recent papers, however, have focused more on applications of charge trap-type flash (CTF) memory, such as semiconductor-oxide-nitride-oxide-silicon, semiconductor-alumina-nitride-oxide-silicon, metal-oxide-nitride-oxide-silicon, and metal-alumina-nitride-oxide-silicon. The reason for this is NAND cell technology is expected to move from floating gate cells to CTF cells beyond the 36-nm node from 2010 [7]. In this review paper, two approaches for TBE are compared: the crested barrier and the VARIOT. Key results of TBE and its applications for NVM are also addressed.

II. Two Approaches for TBE: Crested Barriers and VARIOT

1. Crested Barrier Engineering

In crested barriers, the potential barrier height peaks in the middle and gradually (or abruptly) decreases toward the conducting electrodes. Lihkarev demonstrated through tunneling current simulations that the triple-layer stack using a $Si_3N_4/AlN/Si_3N_4$ combination would enable far better performance than single-layer SiO_2 barriers in terms of programming speed and data retention [4]. These improvements were a result of the high slope of the simulated current voltage of the stack, which is superior compared with one layer of silicon oxide. This can be better understood by comparing the conduction band edge diagrams of Fig. 3a and b [8]. The transparency of a SiO_2 barrier changes slowly with an applied field because the highest part of the barrier, closest to the electron source, is only weakly affected by the applied voltage, as shown in Fig. 3a. In the case of a barrier with a crested conduction band edge profile, the highest part in the middle is pulled very quickly (barrier lowering in Fig. 3b), which allows for transparency and thus the current to change much faster between high and low fields [4]. A critical issue of crested barrier is

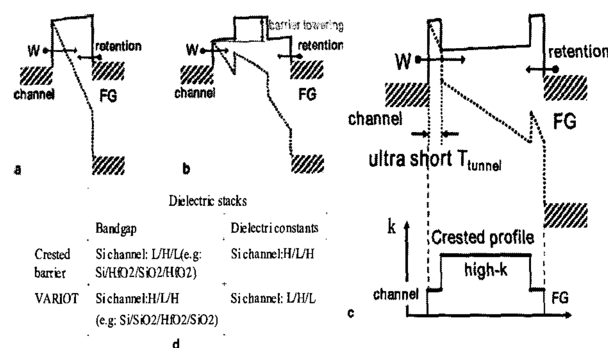


Fig. 3. Conduction band edge diagrams of three different tunnel barriers for low and high fields respectively corresponding to cases of programming and retention. (a) SiO_2 barrier; (b) crested barrier (c-top); VARIOT (c-bottom); crested profile of dielectric constant of VARIOT; (d) summarized table for dielectric stacks in terms of bandgap and dielectric constants [8].

related to the fabrication process. Obtaining good interface quality between the high-k and Si channel is difficult. Only a few reports utilizing crested barrier have been published [9,10].

2. VARIOT Stacks

Govorea et al. proposed a novel engineered tunnel barrier concept called VARIOT, consisting of a two-layer dielectric stack with a low-k/high-k combination (asymmetric) or a three-layer stack with a low-k/high-k/low-k combination (symmetric) with a moderate barrier height of the high layer compared with that of the low layer (i.e., SiO_2) [5]. Note that the crested barrier and the VARIOT have an entirely opposite order in terms of the bandgap and dielectric constant of dielectric stacks, as shown in Fig. 3d. Fig. 4 shows that the VARIOT stack has a much higher field-sensitivity than single SiO_2 tunnel oxide [5]. According to previous research, the maximal acceptable leakage current for a given tunnel barrier, $J_{LEAK} = \Delta Q/(t_{RET}F^2)$, is $\sim 10^{-16}$ A/cm^2 for a square floating gate with an F = 50-nm side length, a total charge Q_{FG} = 200e-, a charge loss ΔQ = 20e-, and a retention time $t_{RET} = 3 \times 10^8$ s (= 10 yr) [8,11]. For voltages $V < V_{RET}$, the current density of the tunnel barrier has to be smaller than J_{LEAK}. On the other hand, for a voltage $V = V_{PGM}$, the tunnel barrier has to be as transparent as possible to program the cells rapidly at moderate electric fields, E (1MV/cm < E < 10 MV/cm) below the breakdown voltages of the materials. The current density corresponding to a t_{PROGR} = 100 μs (1 μs) is on the order of $J = Q_{FG}/(t_{PROGR}F^2) \sim 10^{-2}$ A/cm^2

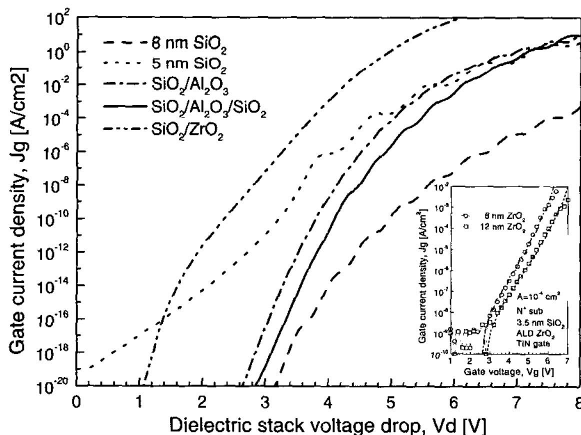
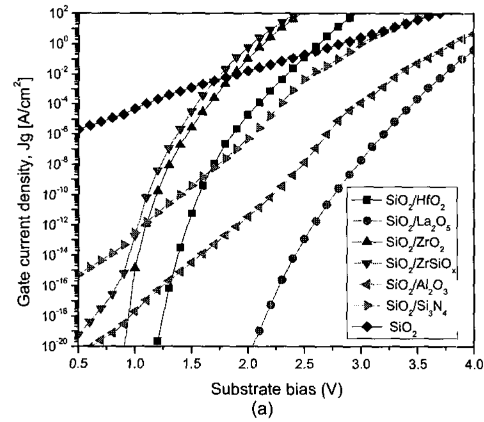


Fig. 4. Current densities versus dielectric stack bias. All stacks have 5-nm EOT [5].



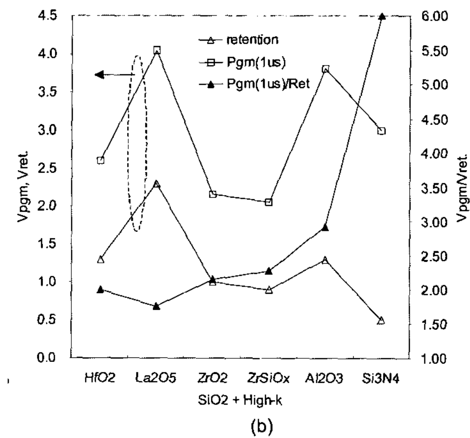


Fig. 5. (a) Gate current densities versus bias for different high-k dielectrics in two-layer asymmetric stacks of 3.0-nm EOT; (b) programming voltage corresponding to 1- s programming, Vpgm, and Vpgm/Vret bias ratio. High-k thickness is [$t_H = (EOT - t_L)\, k_H/k_L$].

(1A/cm^2). To avoid programming disturb, the ratio V_{PGM} / V_{RET} is ideally below 2 [11]. Fig. 5(a) shows a comparison between the tunneling currents of various two-layer dielectric stacks consisting of a SiO_2 and a high-k film with an equivalent oxide thickness (EOT) of 3 nm. It shows that in the case of 3-nm EOT, a single SiO_2 tunnel oxide cannot satisfy the program and retention requirements simultaneously. On the contrary, program/retention performance is improved substantially for SiO_2 and high-k combinations. Figure 5(b) shows that HfO_2 and La_2O_5 provide a promising V_{PGM}/V_{RET} ratio. Among asymmetric (2-layer) and symmetric (3-layer) tunnel barriers, symmetric tunnel barriers are advantageous since they permit similar write and erase times. Figure 6 shows current densities through an asymmetric SiO_2/Al_2O_3 barrier with program and erase bias conditions. Asymmetric barriers allow for

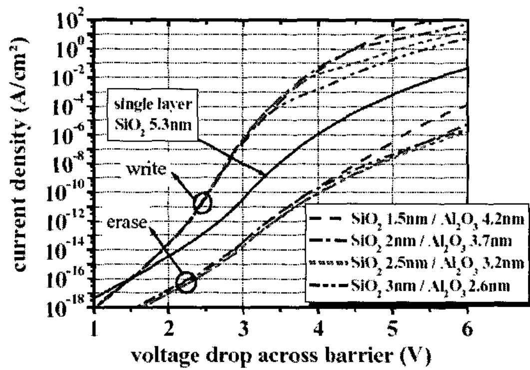


Fig. 6. Current density for asymmetric tunnel dielectrics for program and erase bias conditions. Equivalent oxide thicknesses are 3.3, 3.5, 3.8, and 4.1 nm (from top to bottom) [11].

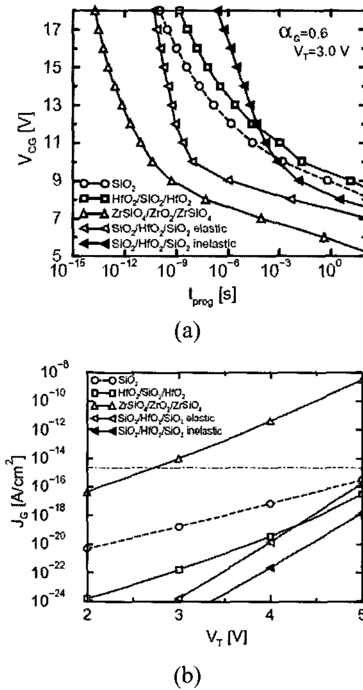


Fig. 7. (a) Control gate voltage (V_{CG}) vs. programming time required to achieve fixed threshold voltage shift; and (b) gate current density due to electrons tunneling out of floating gate electrode during data retention at V_{CG} = 0V as a function achieved during programming [12].

even steeper I-V slopes (= faster programming in "write" mode than that in the symmetric barriers). However, in "erase" mode the required voltages are higher than those for SiO_2, as shown in Fig. 6. Therefore, erase mechanisms other than standard FN tunneling, such as hot carrier injection, may be needed to fully exploit the advantages of this asymmetric tunnel barrier in a tradeoff with power consumption.

The crested barrier and the VAROT approaches have both been proven effective for enhancing field sensitivity. The question is which is more effective? Driussi and Buckley presented a theoretical analysis of both crested barriers and VARIOT with regard to NVM devices [12,13]. The main conclusion was that the VARIOT (low-k/high-k/low-k or low-k/high-k) dielectric stack combination is a more appropriate candidate as a tunnel barrier for NVM. As shown in Fig. 7a, the VARIOT stack (the HLH barrier from the point of view of the energy band or the LHL barrier from the point of view of the dielectric constant) can potentially achieve smaller programming times and/or smaller programming voltages than single oxide and the crested barrier (on the assumption that charge transport in the high-k layer is not entirely inelastic [12]. Figure 7b shows that VARIOT also has a better retention time than single oxide and the crested barrier [12].

III. Band Engineering for CTF Memory

At an early stage, TBE was mostly applied to floating gate flash memory [14,15]. Since there is an urgent need for CTF devices to have flash memory scaling beyond the 36-nm node [7], recent research is more

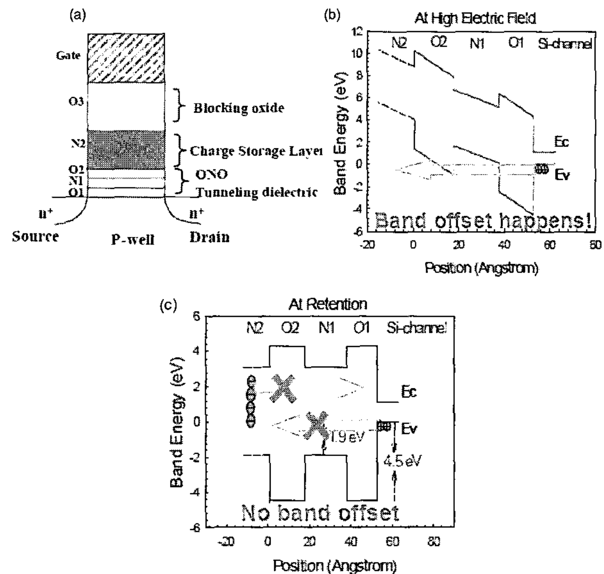


Fig. 8. (a) Structure of n-channel BE-SONOS using ultra-thin ONO tunneling dielectric. Bandgap engineering concept for BE-SONOS device. (b) At high field during erase, band offset reduces hole-tunneling barrier to merely O1. (b) At low field during retention, both electron de-trapping and hole back tunneling are suppressed by full O1/N1/O2 barrier stack [16].

focused on CTF memory. Hang-Ting Lue et al., proposed bandgap engineered SONOS (BE-SONOS) [16]. In BE-SONOS, the ultra-thin "O/N/O" provides a "modulated tunneling barrier", as shown in Fig. 8. The principal benefit comes from the low hole-tunneling barrier (~1.9 eV) of SiN compared with that of oxide (~4.5 eV). Under a low electrical field during retention, both electron de-trapping and hole back tunneling cannot take place because the total thickness of the O1/N1/O2 barrier is large. On the other hand, under a high electric field during erase, band offset occurs and the hole barriers of N1 and O2 are below the valence band, thereby reducing the effective hole-tunneling distance to the thin O1. As a result, a large hole-tunneling current through O1 occurs under –FN bias. Hang-Ting Lue et al., obtained a self-convergent erased Vt (threshold voltage) of ~3 V, suitable for a NOR application with an N+ poly gate, and a depletion mode device (Vt < 0) for a NAND application with a P+ poly gate. Sheng-Chih Lai et al. reported on different erase mechanisms for MANOS, MONOS, and BE-SONOS (Fig. 9) [17]. The group also reported that BE-SONOS has the fastest erase speed and a better data retention than both MANOS and MONOS. They also reported a novel BE-SONOS with a gate injection for program and

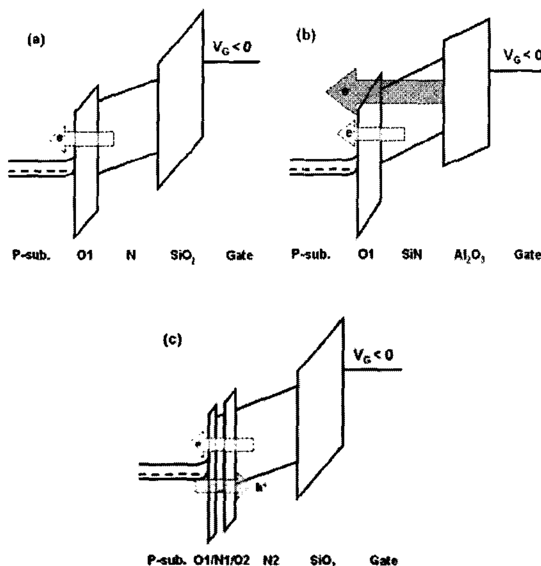


Fig. 9. Schematic diagram illustrating erase mechanism of (a) MONOS, (b) MANOS, and (c) BE-SONOS. For MONOS, erase comes from electron de-trapping from nitride traps. For MANOS, additional shallower traps contribute to enhanced erase speed. For BE-SONOS, erase is mainly a result of hole tunneling from Si valence band [17].

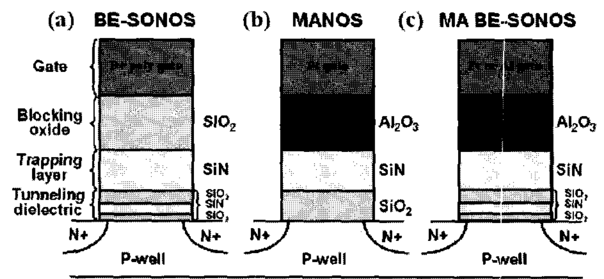


Fig. 10. Structures of (a) BE-SONOS, (b) MANOS, and (c) MA BE-SONOS devices [18].

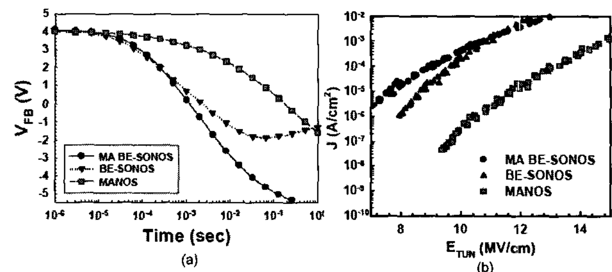


Fig. 11. (a) Erase curves (V_{FB}-time) for three devices (BE-SONOS, MANOS, and MA BE-SONOS) at VG = -18 V, where gate material for MA BE-SONOS and MANOS is platinum [18]. (a) J-E plots using transient analysis [18].

erase (P/E) operations [18] and BE-SONOS using an Al_2O_3 top blocking layer and a metal gate (MA BE-SONOS) [19], as shown in Fig. 10, to provide a very fast erase speed without erase saturation. For gate injection BE-SONOS, the device used is a p-channel BE-SONOS having ultra-thin ONO tunneling dielectric grown on top of the trapping nitride. Programming and erasing are done by –FN electron injection and +FN hole injection from the poly gate, not from the Si channel side. Very high endurance (10-M cycle P/E) was reported to be a result of suppressed channel interface degradation. For MA BE-SONOS, a faster erase speed was achieved, owing to a bandgap engineered ONO barrier that facilitates hole tunneling, compared with MANOS using a thick (4.5 nm) tunnel oxide. Unlike BE-SONOS using a P+-poly gate and top oxide, MA BE-SONOS did not show erase saturation, owing to the metal gate and Al_2O_3 blocking layer, which greatly reduce gate injection during erase, as shown in Fig. 11. Furnémont et al. reported that both read and program disturb sensitivity are major drawbacks for nitride-based NAND flash and that VARIOT has a much better disturb sensitivity than conventional SANOS memory, as shown in Fig. 12 [20].

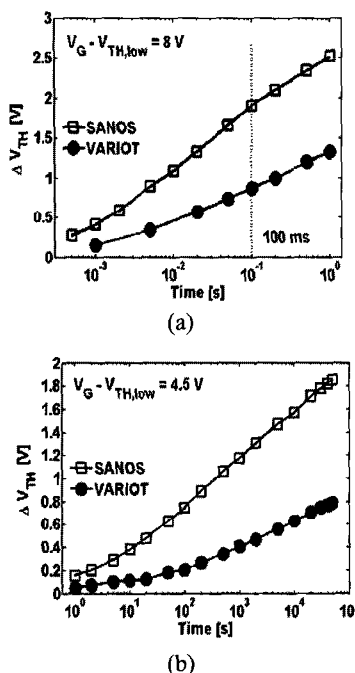


Fig. 12. Measurement comparing (a) program disturb, and (b) read disturb between SANOS and VARIOT. VARIOT case shows less than half the disturb that SANOS has, and more than two decades of disturb time are gained with VARIOT for same V_{TH} shift.

Band engineering can be applied not only to tunneling dielectrics but also to other areas, such as charge trap layers. ZongLiang Huo et al. were first to demonstrate multi-level CTF devices with a band-engineering concept used for the trap layer [21]. The engineered band structure, $Si_3N_4/Al_2O_3/Si_3N_4$ (NAN), was used as a trap layer in place of a single Si_3N_4 layer in the MANOS structure. For this structure, the charge centroid position moves toward the top blocking layer while filling the nitride with electrons. High electron density at the Si_3N_4/blocking oxide interface increases the leakage current. By inserting an Al_2O_3 layer into the Si_3N_4, electron distribution becomes flatter, and a lower leakage current can be obtained as a result of the low electron density near the blocking layer. Thus, for the same memory window and P/E time, a low program voltage/low gate current can be obtained for this novel structure.

IV. CONCLUSIONS

Two approaches for TBE were compared: the crested barrier and the VARIOT. The VARIOT stack combination is a better candidate as a tunnel barrier for NVM. In its early stages, tunnel barrier engineering was mostly applied to floating gate flash memory. Since an urgent need has arisen for CTF devices to have flash memory scaling beyond the 36-nm node, recent research is more focused on CTF memory. This memory with TBE shows a much better disturb sensitivity than that of conventional SANOS memory. BE-SONOS or MA BE-SONOS, which uses an ultra-thin ONO tunnel barrier, has potential for future CTF memory. They exhibit substantially faster erase speeds than those of conventional SONOS, owing to a bandgap engineered ONO barrier that facilitates hole tunneling. In addition, band engineering of charge trap layers and the tunneling barrier is an important issue that will need to be addressed for future CFT memory.

REFERENCES

[1] S. Lai, "Tunnel oxide and ETOX™ flash scaling limitation," *1998 Int'l Non-Volatile Memory Technology Conference*, pp. 6-7, 1998.

[2] K. Naruke et al., "Stress induced leakage current limiting to scale down EEPROM tunnel oxide thickness," *IEDM Tech. Dig.*, pp. 424-427, 1988.

[3] "Front end processes," in International Technology Roadmap for Semiconductors 2003 Edition. Austin, TX: Semiconductor Industry Assoc., 2003.

[4] K. K. Lihkarev, "Layered tunnel barriers for nonvolatile memory devices," *Appl. Phys. Lett.*, vol. 73, no. 15, pp. 2137-2139, Oct. 1998.

[5] B. Govoreanu, P. Blomme, M. Rosmeulen, J. Van Houdt, and K. De Meyer, "VARIOT: a novel multilayer tunnel barrier concept for low-voltage nonvolatile memory devices," *IEEE Electron Device Lett.*, vol. 24, no. 2, pp. 99-101, Feb. 2003.

[6] "Emerging research devices," in International Technology Roadmap for Semiconductors 2003 Edition. Austin, TX: Semiconductor Industry Assoc., 2003.

[7] "Emerging research devices," in International Technology Roadmap for Semiconductors, Winter Conference, Makuhari, Japan, 2007.

[8] J. Buckley, B. De Salvo, G. Ghibaudo, M. Gely, J. F. Damlencourt, F. Martin, G. Nicotra, and S.

Deleonibus, "Investigation of SiO_2/HfO_2 gate stacks for application to non-volatile memory devices," *Solid-State Elect.*, vol. 49, pp. 1833-1840, 2005.

[9] Seung Jae Baik, Siyoung Choi, U-In Chung, and Joo Tae Moon, "Engineering on tunnel barrier and dot surface in Si nanocrystal memories," *Solid-State Elect.*, vol. 48, pp. 1475-1481, 2004.

[10] Julie D. Casperson, L. Douglas Bell, and Harry A. Atwater, "Materials issues for layered tunnel barrier structures," *J. Appl. Phys.*, vol. 92, no. 1, pp. 261-267, 2002.

[11] M. Specht, M. Städele, and F. Hofmann, "Simulation of high-K tunnel barriers for nonvolatile floating gate memories," *Proc. ESSDERC Conference*, pp. 599-602, 2002.

[12] F. Driussi, S. Marcuzzi, P. Palestri, and L. Selmi, "Gate current in stacked dielectrics for advanced FLASH EEPROM cells", *Proceedings of ESSDERC*, Grenoble, France, pp. 317-320, 2005.

[13] J. Buckley et al., "Engineering of conduction band crested barriers or dielectric constant crested barriers in view of their application to floating-gate non-volatile memory devices," *Silicon Nanoelectronics Workshop*, 2004.

[14] Y. Liu, S. Dey, S. Tang, D. Q. Kelly, J. Sarkar, and S. K. Banerjee, "Improved performance of SiGe nanocrystal memory with VARIOT tunnel barrier," *IEEE Trans. Electron. Devices*, vol. 53, no. 10, pp. 2598-2602, 2006.

[15] P. Blomme, J. D. Vosa, A. Akheyar, L. Haspeslagha, J. V. Houdt, and K. D. Meyer, "Scalable floating gate flash memory cell with engineered tunnel dielectric and high-K (Al_2O_3) interpoly dielectric," *IEEE Non-Volatile Semiconductor Memory Workshop*, pp. 52-53, 2006.

[16] Hang-Ting Lue, Szu-Yu Wang, Erh-Kun Lai, Yen-Hao Shih, Sheng-Chih Lai, and Ling-Wu Yang, "BE-SONOS: a bandgap engineered SONOS with excellent performance and reliability," *IEDM Tech. Dig.*, pp. 547-550, 2005.

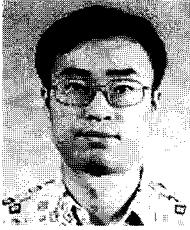
[17] Sheng-Chih Lai et al., "Study on the erase and retention mechanisms for MONOS, MANOS, and BE-SONOS non-volatile memory devices," *Symp. on VLSI Tech. Digest of Technical Papers*, pp. 1-2, 2007.

[18] Hang-Ting Lue et al., "A novel gate-injection program/erase p-channel NAND-type flash memory with high (10M cycle) endurance," *Symp. on VLSI Tech. Digest of Technical Papers*, pp. 140-141, 2007.

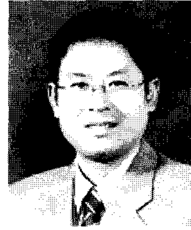
[19] Sheng-Chih Lai, et al., "MA BE-SONOS: a bandgap engineered SONOS using metal gate and Al_2O_3 blocking layer to overcome erase saturation," *Non-Volatile Semiconductor Memory Workshop*, pp. 88-89, 2007.

[20] A. Furnémont, M. Rosmeulen, A. Cacciato, L. Breuil, K. De Meyer, H. Maes, and J. Van Houdt, "Physical understanding of SANOS disturbs and VARIOT engineered barrier as a solution," *Non-Volatile Semiconductor Memory Workshop*, pp. 94-95, 2007.

[21] ZongLiang Huo, JunKyu Yang, SeungHyun Lim, SeungJae Baik, Juyul Lee, JeongHee Han, In-Seok Yeo, U-In Chung, Joo Tae Moon, and Byung-Il Ryu, "Band engineered charge trap layer for highly reliable MLC flash memory," *2007 Symposium on VLSI Technology Digest of Technical Papers*, pp. 138-139, 2007.



Jongwan Jung received the M.S. and Ph.D. degrees in electrical engineering from the Korea Advanced Institute of Science and Technology, Seoul, Korea, in 1991 and 1996, respectively. From 1996 to 2001, he was a Senior Member of Technical Staff in the field of logic and DRAM device development, at Hynix Semiconductor, Inc., Korea. From 2001 to 2004, he was a Postdoctoral Researcher at the Massachusetts Institute of Technology, Cambridge, MA. During that period, he researched SOI and SiGe-Si devices. He received the IEEE Electron Devices Society George E. Smith in 2004. From 2004, to 2006, he was with Samsung Electronics, developing CMOS image sensors with sub-2 m pixels. He is now an associate professor in the department of nano science and technology at Sejong university. His current interests include the design and fabrication of nanoscale CMOS, non-volatile memories, CMOS image sensors, thin film transistors and graphene transistors.



Won-Ju Cho received the B.S. degree from Kyungpook National University, Daegu, Korea, in 1989, and the M.S. and Ph.D. degrees in electrical engineering from Keio University, Tokyo, Japan, in 1991 and 1994, respectively. During 1994-2000, he joined in the HYNIX Semiconductor Inc., Korea. During 1999-2000, he was with the AIST, Tsukuba, Japan, where he developed the ultra-low-loss power SiC devices. During 2000-2004, he joined in Electronics and Telecommunications Research Institute (ETRI), Daejon, Korea, for development of nanoscale silicon MOSFET devices and nano-bio sensors. Recently, he was appointed Professor of Department of Electronic Materials Engineering at the Kwangwoon University, Seoul, Korea. His current research interests include new electronic materials and devices, simulations, processing, and analysis of nanoscale silicon devices, non-volatile memories, Bio-devices, micro- & nano-electromechanical systems (MEMS & NEMS).