A Novel Self-Aligned 4-Bit SONOS-Type Nonvolatile Memory Cell With T-Gate and I-Shaped FinFET Structure

Sunyeong Lee, *Student Member, IEEE*, Yong Woo Jeon, Tsu-Jae King Liu, *Fellow, IEEE*, Dae Hwan Kim, *Member, IEEE*, and Dong Myong Kim, *Member, IEEE*

***Abstract*—We propose a novel 4-bit self-aligned SONOS-type nonvolatile memory (NVM) cell with a T-gate and I-shaped FinFET structure for practical implementation with high storage density and better reliability. In order to obtain enhanced reliability characteristics, a modified Fowler–Nordheim tunneling mechanism is employed for programming along the channel length direction, while a band-to-band hot hole injection is used for erasing along the channel width direction. With separated paths for program and erase, improved device performance is obtained with sensing margin. In order to improve the immunity to second-bit effects, the gate-induced drain leakage current method, which is a charge detection method and highly sensitive to the locally stored charges, is employed for the reading of the stored data. In terms of the scalability, we confirmed by 2-D technology computer-aided design simulation that the proposed NVM cell with channel length $L = 50$ nm operates with enough sensing margin and high-density ($\sim 5\ \mathrm{F}^2/\mathrm{bit}$) NVM by the crossed cell array architecture.**

***Index Terms*—Double gate, field-effect transistor, fin field-effect transistor (FinFET), Fowler–Nordheim (F-N) tunneling, gate-induced drain leakage (GIDL), multibit memory, nonvolatile memory (NVM), silicon-oxide-nitride-oxide-silicon (SONOS).**

I. Introduction

CONVENTIONAL planar Flash memories are based on the polysilicon floating gate as a charge storage location. Therefore, the conductive polysilicon (poly-Si) floating gate as a charge storage medium is subject to the stress-induced leakage current (SILC) caused by accumulated traps in the insulating oxide (SiO_2) during the program/erase (P/E) cycling. The SILC may cause charges to leak away from the floating gate resulting in the volatility of the memory cell and a degradation of the retention characteristics. Consequently, the scaling of the oxide thickness down to below 8 nm is critically restricted due to a significant loss of stored charges after repeated P/E cycling [1]. In addition, the equivalent oxide thickness (EOT) of the thick gate stack makes it difficult to scale the gate length down to below 50 nm [2].

The silicon-oxide-nitride-oxide-silicon (SONOS)-type Flash memory employing a nonconductive nitride (Si_3N_4) trapping layer instead of the conducting poly-Si floating-gate storage element is expected to be the most promising for a nonvolatile memory (NVM) device. SONOS-type Flash memory devices are more scalable than floating-gate memory devices because the charges are stored in discrete traps within the Si_3N_4 layer allowing a thinner tunnel oxide to be used. This makes SONOS-type devices more robust to the SILC phenomenon and triggers the SONOS structure to be developed as one of the next-generation Flash devices [3], [4]. We also note that the SONOS-type NVM is free from the floating-gate coupling interference between storage cells [5], and it is under active study as a promising high-density Flash memory technology [6].

For further enhancement of the storage density, multiple bits can be implemented in a single NVM cell. A dual-bit charge storage scheme, based on charge trapping at different locations in a single charge-trapping layer, is currently used in conventional single-gate SONOS NVM structures. For this purpose, the NROM [7] introduced two distinct mechanisms: the channel hot electron injection (CHEI) for program and the band-to-band tunneling-induced hot hole for erase. This leads to the injection of charges in a specific localized area in the Si_3N_4 storage layer. The nonconductive Si_3N_4 layer makes the NROM store two distinct bits in a single cell, although there are disadvantages, including a long physical distance (the second-bit effects [8]) and high drain bias (V_{DS}) for the CHEI during programming. In order to provide a superb electrostatic integrity, a multiple-gate transistor structure such as a fin field-effect transistor (FinFET) [9] has been also reported even with a thick EOT as far as the body is sufficiently thin (i.e., much thinner than the gate length L_g) [10]. To shorten the gate length down to 20 nm [11, 12], multiple-gate SONOS NVM cells have been investigated as a single-bit storage cell [10], a multilevel charge storage cell [11], and a dual-bit storage cell [13].

Although multiple-gate cells have been employed in SONOS NVMs, charges are still stored at physically distinct locations in the Si_3N_4 layer causing interference (as a second-bit effect) between neighboring bits. Recently, a novel multibit NVM, based on the physically separated charge storage such as the gate-sidewall storage [14]–[19], has been proposed to improve the mismatch of the injected electrons and holes, the lateral drift

Manuscript received January 11, 2010; accepted April 21, 2010. Date of publication May 24, 2010; date of current version July 23, 2010. This work was supported by Kookmin University Research Initiative. The CAD software was supported by SILVACO and the IC Design Education Center. The review of this paper was arranged by Editor G.-T. Jeong.

S. Lee, Y. W. Jeon, D. H. Kim, and D. M. Kim are with the School of Electrical Engineering, Kookmin University, 861-1 Jeongneung-Dong, Seongbuk-Gu, Seoul, 136-702, Korea (e-mail: dmkim@kookmin.ac.kr).

T.-J. K. Liu is with the Department of Electrical Engineering and Computer Sciences, University of California, Berkeley, CA 94720-1776 USA.

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.2010.2049672

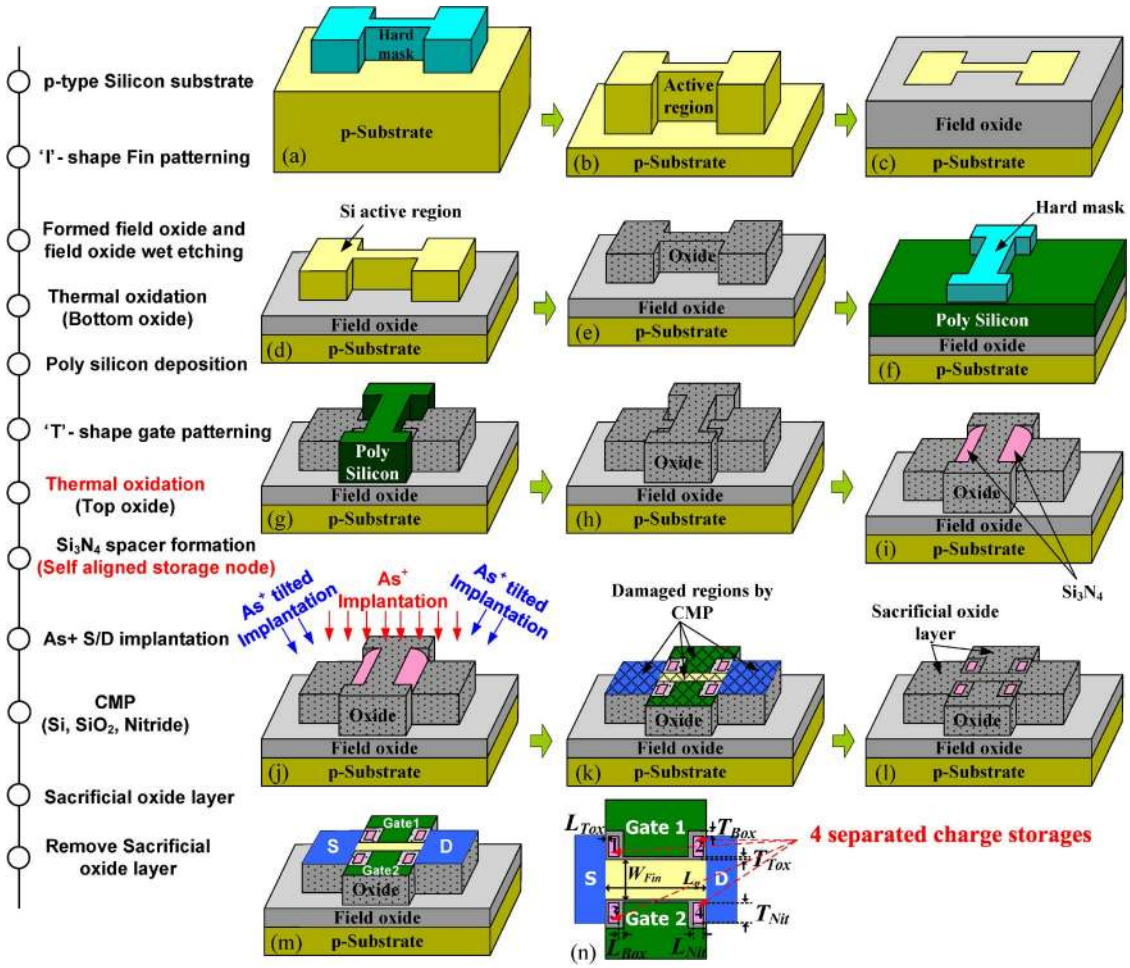

Fig. 1. Fabrication process flow for the proposed 4-bit TGIF NVM cell. (a) I-shaped active fin patterning. (b) Field oxide formation. (c) Tunnel oxide formation by thermal oxidation. (d) Polysilicon gate electrode formation. (e) Blocking oxide formation by thermal oxidation of the "T"-shaped polysilicon gate. (f) Filling of the four holes between the blocking and tunnel oxides by Si_3N_4 deposition and etch-back. (g) Separation of the front- and back-gates from Si_3N_4 charge-storage regions by CMP. (h) Four-bit TGIF NVM cell structure with design parameters: $L_g = 100$ nm, $W_{\rm fin} = 40$ nm, $T_{\rm TOX} = L_{\rm TOX} = 3$ nm, $T_{\rm Nit} = 10$ nm, $L_{\rm Nit} = 15$ nm, $T_{\rm Cox} = L_{\rm Cox} = 5$ nm, $N_{\rm Sub} = 1 \times 10^{16}$ cm^{-2}, $N_{\rm poly} = N_{\rm SD} = 1 \times 10^{20}$ cm^{-3}, and S/D doping gradient ≈ 1 nm/dec.

of stored charges, the recombination of remaining charges after the P/E cycling [20], [21], and the second-bit effects. However, these structures still have problems to be solved, i.e., the stored charges next to the drain electrode affect the threshold voltage V_T of the cell when the gate length is scaled down to well below 50 nm. To suppress these effects, a novel read method sensed by the band-to-band tunneling (BTBT) current enhanced by the locally stored charges [22] is considerable.

In this paper, we propose a novel self-aligned T-gate and I-shaped FinFET (TGIF) SONOS-type NVM cell with physically separated 4-bit storages and verified the performance by 2-D technology computer-aided design (TCAD) simulation. The gate-induced drain leakage (GIDL) current read method is employed for more sensitive detection of the locally stored charges within the gate-dielectric stack. Since less stored charge is needed in the sensitive GIDL read method, deep scaling of the gate length and operation at low P/E voltages is possible resulting in improved endurance characteristics [23]. For better reliability characteristics, we used a novel program/erase scheme employing separated paths for program from the path for erase. In order to improve the memory density, crossed memory cell array architecture is also suggested.

II. Device Structure, Array, and Fabrication Process for the TGIF SONOS-Type NVM Cells

The proposed 4-bit self-aligned TGIF SONOS NVM cell structure is schematically shown in Fig. 1(n) with stepwise fabrication process flow. As one of the key features, the T-shape gate electrode can be divided into two sectional parts: 1) an intergate and 2) a normal gate. The intergate electrode is used for lateral programming (*in the channel length direction*) through the gate-to-source/drain electrode overlap regions. Furthermore, the gate controllability is improved by the intergate electrode employing a thin EOT. On the other hand, the normal gate above the ONO stack contributes to the enhancement of the vertical field (*in the channel width direction*) for the BTBT process during the erase step. The detailed program and erase schemes will be described in Section III. The I-shape active fin region is another important part of the proposed 4-bit TGIF SONOS device structure. The overlapped (raised) source/drain electrodes enable the charges to transfer through the lateral paths with the intergate electrode under high lateral electric fields.

In the simulation work for the verification, we build a 4-bit TGIF SONOS device with a thin-body n-channel field-effect transistor having channel length $L = 100$ nm and multilayered

gate dielectrics. It includes a charge-trapping layer (Si_3N_4 layer with length = 10 nm and thickness = 15 nm) sandwiched between the thin tunnel oxide layer (3.8 nm; SiO_2) and the thick control oxide layer (5 nm; SiO_2). The *length* of each Si_3N_4 storage node is determined by the effective area for the BTBT process during the erase, and it depends on the source/drain junction edges and doping profiles. On the other hand, the *thickness* of each Si_3N_4 storage node is determined by the program/erase efficiency as well as by the effective area for the BTBT process. If the Si_3N_4 storage node is not thick enough to reduce the vertical electric field in the channel width direction, the unselected bit is subject to be programmed/erased by the high vertical electric field. We note that the tunnel oxide ought to be thin enough, allowing electrons to be effectively injected from the channel layer into the charge-trapping region during the program operation. On the other hand, the control oxide should be thick enough to suppress a possible injection of electrons from the gate electrode into the charge-trapping Si_3N_4 layer during the erase operation. The lightly doped p-type silicon channel/body is employed to improve the carrier mobility and to reduce the random dopant fluctuation effect [24]. In order to suppress the short-channel effects (SCEs), the fin body width is set to be $W_{\rm fin} \leq 0.45 \times L_g$ as recommended [25].

The device structure and possible process flow for practical implementation of the proposed 4-bit TGIF SONOS NVM cells are schematically illustrated in Fig. 1. The first step is a patterning of the I-shaped active region on the bulk wafer [Fig. 1(a) and (b)], followed by the formation of the field oxide [Fig. 1(c) and (d)] for isolation. The tunnel oxide is then grown on the surface of the exposed silicon active fin [Fig. 1(e)]. A T-shaped polysilicon gate electrode is formed [Fig. 1(f) and (g)], and then, the blocking oxide layer is thermally grown on the surface of the gate [Fig. 1(h)]. We note that the final thickness of the tunnel oxide is confirmed to be 8 Å even after the thermal growth of the blocking oxide [26]. Contrary to the conventional SONOS NVMs, which use a deposited block oxide above the Si_3N_4 trapping layer, the proposed 4-bit TGIF SONOS NVM uses a thermal oxidation for better quality and reliability characteristics of the oxide layer. From the simulation results, we also confirmed that 20% misalignment of the gate electrode is still good for a proper operation of the proposed TGIF SONOS NVM cells.

As a key fabrication step to form the charge storage nodes by the self-alignment process, four holes between the blocking oxide and the tunnel oxide are filled by conformal deposition and successive anisotropic etch-back of the Si_3N_4 layer [Fig. 1(i)]. In conventional SONOS NVMs, on the other hand, it is complicated to build physically separated storage nodes with extra masks. We also note that it is advantageous to replace the Si_3N_4 layer with other materials such as metals with high workfunction, as usually adopted in nanocrystal memory structures. We can make a 4-bit NVM cell with a long retention characteristic utilizing a large difference in the workfunction between the metal and the oxide [27], [28]. After source/drain doping (n-type) by ion implantation and successive annealing [Fig. 1(j)], the front gate, back gate, and Si_3N_4 storage nodes are separated by the chemical mechanical polishing (CMP) process to allow independent operation of each storage node [Fig. 1(k)]. Damage caused by the CMP process can be cured by a sacrificial oxidation process [Fig. 1(l)].

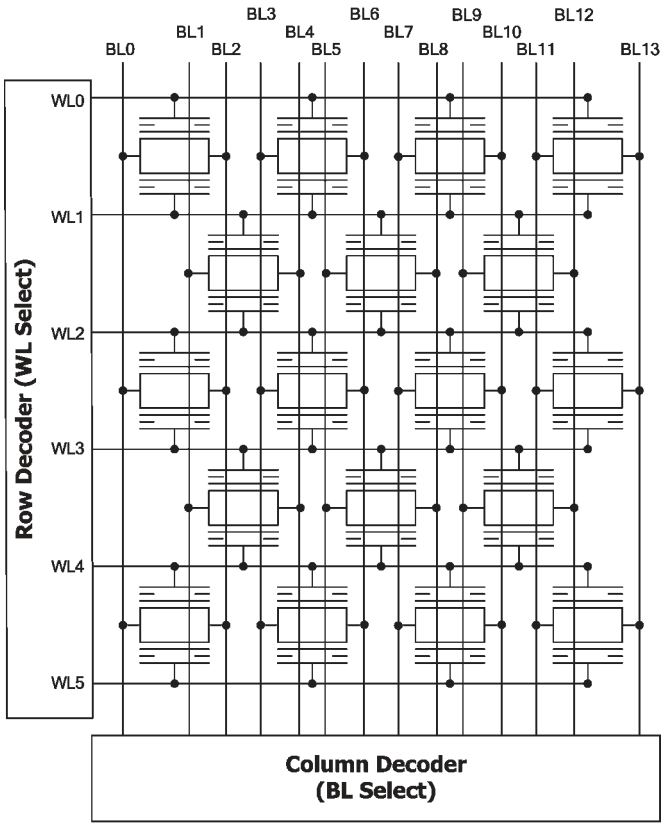

Fig. 2. Schematic of the NOR-type 4 × 5 TGIF NVM cell array.

The 4-bit TGIF SONOS NVM can be highly dense ($\sim$5 F^2/bit) when the cells are placed in crossing arrays, as shown in Fig. 2. In order to further increase the memory density, the gate2 electrode of each row is designed to share a word line with the gate1 electrodes of the next one. If the gate electrodes are connected not by the metallic word line but by the polysilicon word line, the cells can be closer because a metal contact is not necessary for each gate. In the case of the bit line, the source/drain electrodes of the first row are connected to the third-row ones for low parasitic capacitances of the bit line and low leakage current from unselected cells on the bit line.

III. GIDL Current Read Method and Separated Paths for Program/Erase

A. GIDL Current Read Method

Based on the program or the erase state of four storage nodes, 16 binary states are defined in the proposed 4-bit TGIF NVM cells. The binary state is defined as "0000" when four nodes are erased, while it is defined as "1111" when they are all filled with programmed electrons. For instance, if a storage node (Bit1) sandwiched between the source and the gate1 is only programmed, we define the binary state as "1000." Similarly, if a node (Bit4) sandwiched between the drain and the gate2 is programmed, its binary state is defined as "0001." The 16 binary states can be read by the cell's GIDL current change with the stored charges near the drain electrode. As illustrated

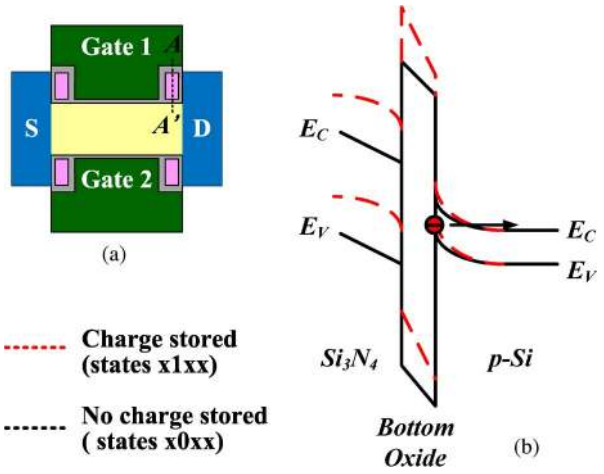

Fig. 3. Schematic of the GIDL current read method. (a) Cut view of the structure. (b) Energy band diagram illustrating the effect of charges stored on Bit2 (close to the drain) on the cell's GIDL current. The programmed charges near the drain electrode (red dotted line) enhance the transverse electric field, which results in a larger GIDL current due to the band-to-band tunneling.

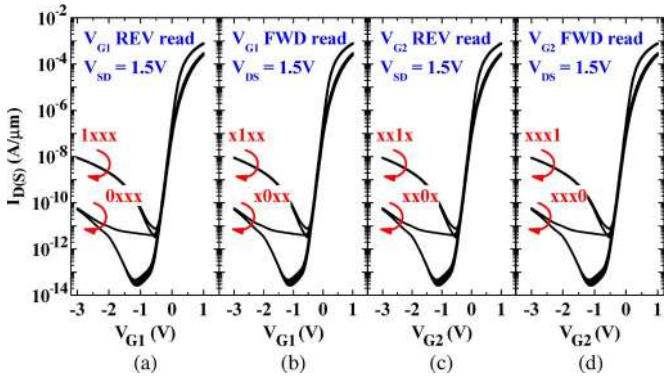

Fig. 4. $I_{D(S)}-V_{G(1,2)}$ characteristics of the 4-bit TGIF SONOS-type memory device. (a) Sensing of Bit1 by reverse reading ($V_{SD} = 1.5$ V) of the GIDL current under negative V_{G1}. (b) Sensing of Bit2 by forward reading ($V_{DS} = 1.5$ V) of the GIDL current under negative V_{G1}. (c) Sensing of Bit3 by reverse reading ($V_{SD} = 1.5$ V) of the GIDL current under negative V_{G2}. (d) Sensing of Bit4 by forward reading ($V_{DS} = 1.5$ V) of the GIDL current under negative V_{G2}.

in Fig. 3, the programmed charges near the drain electrode enhance the transverse electric field resulting in enhanced GIDL current due to the BTBT process [22], [23].

As summarized in Fig. 4, the validity of the GIDL current read method was confirmed by the 2-D TCAD device simulation for a TGIF SONOS NVM cell with $L_g = 100$ nm and $W_{\rm Fin} = 40$ nm. Advanced physical models are considered with quantum effects, tunneling mechanisms, and mobility models [26]. In the simulation, obtained from the program characteristics as shown in Fig. 5, the volume density of charges stored in the programmed charge-trap node was set to 2×10^{19} C/cm^3 (uniform space and energy charge profile). When Bit2 is programmed (States "x1xx"), as can be seen in Fig. 4(b), the GIDL current was enhanced so that the OFF-state current in the transistor cell can be used to identify the state of Bit2 by applying a negative gate1 voltage ($V_{G1S} \cong -3$ V) under a moderate drain-to-source voltage (e.g., $V_{DS} = 1.5$ V; a high OFF-state current $\rightarrow$ electrons are stored in the storage node). We note that gate2 is negatively biased to reduce the

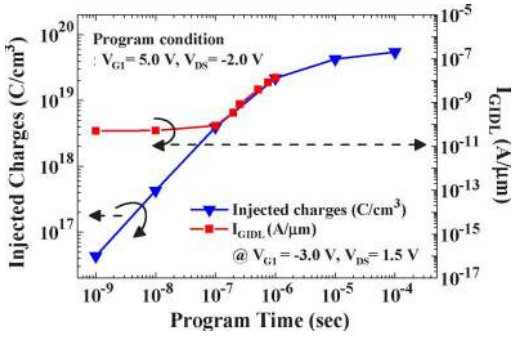

Fig. 5. Transient simulation results for the 4-bit TGIF SONOS NVM with ▼ = injected charges and ■ = $I_{\rm GIDL}$ induced by the injected charges.

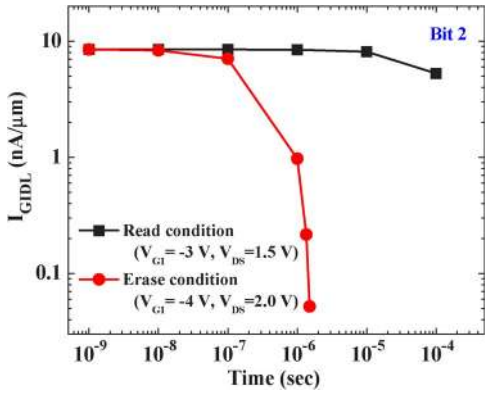

Fig. 6. Simulation results of $I_{\rm GIDL}$ transient characteristics with ■ = the read condition and ● = erase condition.

subthreshold leakage current caused by low V_T. If gate2 is biased at a high negative voltage, the BTBT current generated by Bit4 also increases (read disturbance). It is worthy to note that the read condition ($V_{G1S} \cong -3$ V, $V_{DS} = 1.5$ V) should be low enough not to erase the stored charges during the read operation, as shown in Fig. 6. In order to determine the logic state of Bit4 (located on the opposite side of Bit2), a negative bias on gate2, a positive bias on the drain, and a slightly negative bias on gate1 need to be applied. The difference in the OFF-state currents between the programmed and erased states is large enough ($> 10^2$ A/μm) to determine the logic state of each bit.

When the gate electrode is misaligned, the sensing margin keeps at $\sim 10^2$ A/μm, and the sensed GIDL current is not high enough even on the programmed state. Therefore, a modified boosting current amplifier is employed to increase the sensing current and maximize the on/off ratio [29]. As a result, the sensing current from a programmed cell can be increased up to $\sim 10^{-4}$ A/μm from $\sim 10^{-8}$ A/μm, and the GIDL current of the erased cell is depressed at least ($\sim 10^{-12}$ A/μm). Finally, by using the modified current amplifier, we obtained a large on/off current ratio ($\sim 10^8$ A/μm) (Fig. 7) for sensing the charge state with the GIDL current read method.

We also note that the OFF-state GIDL current is robust to the logic state of neighboring bits. This result indicates that the GIDL read method is less subject to the SCE (and, hence, to the second-bit effects) than the conventional V_T read method, and so it can be used in NVM cells with extremely scaled gate

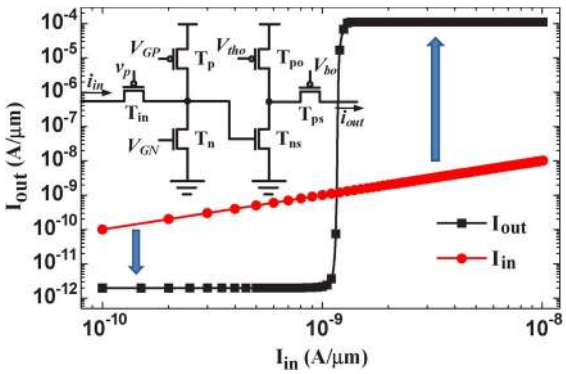

Fig. 7. Simulation result (● = input current and ■ = output current) for the current amplifier with an inset for the circuit.

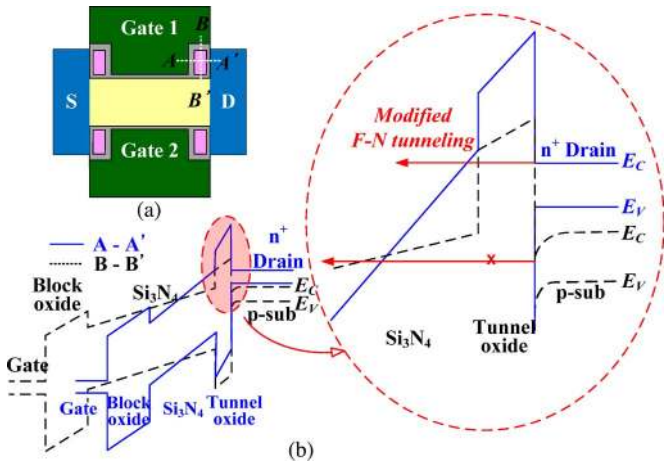

Fig. 8. Schematic energy band diagram under a modified F-N tunneling programming condition. A-A′: along the gate-to-drain lateral direction (channel length direction); B-B′: across the gate-to-substrate vertical direction (channel width direction).

lengths. Moreover, the GIDL read method is more sensitive to the stored charges than the conventional V_T read method. As can be seen in Fig. 4, it is difficult to determine the stored charge state by the V_T read method ($I_D = 1 \times 10^{-6}$ A/μm) in contrast to the GIDL read method ($V_G = -3$ V). Due to the small amount of charges for sensing the charge state in the GIDL read method, program/erase can be performed at reduced voltage resulting in shorter program/erase time and, thus, improved reliability characteristics.

B. Separated F-N Tunneling Paths for the Program/Erase and Simulation Result

It is important to note that the only selected bit is programmed (or erased), while other bits are not disturbed during the program (or erase) operation. Therefore, in order to inject charges only into a selected bit for programming, we used modified Fowler–Nordheim (F-N) tunneling [30] through the gate-to-drain (or gate-to-source) overlap region with positive bias at gate1 for Bit1 and Bit2 (or gate2 for Bit3 and Bit4), while negative bias at the drain for Bit2 and Bit4 (or source for Bit1 and Bit3). The energy band diagram for the Bit2 programming condition is schematically shown in Fig. 8. For enhanced lateral electric field while suppressing the vertical field for effective programming, a negative bias is applied to the drain electrode with thick charge storage nodes. In this bias condition, the applied gate-to-drain bias is not high enough to bend the energy band for F-N tunneling through the drain and the tunnel oxide barrier. However, for modified F-N tunneling through the drain and Si_3N_4 barrier, the program condition is high enough to bend the energy band resulting in a *lateral* program path for fast programming. For erasing the selected bit, we use a band-to-band hot hole injection (BBHHI), which is a common method to erase multibit NVMs. We also checked the band diagram and the Shockley–Read–Hall generation rate by the 2-D TCAD simulation under erase conditions [26]. With separated paths for program and erase, improved device performance is obtained with sensing margin. In order to improve the immunity to second-bit effects, the GIDL current method, which is a charge detection method and highly sensitive to the locally stored charges, is employed for the reading of the stored data.

We also note that program and erase conditions are very crucial to the reliability characteristics of the proposed 4-bit TGIF NVM cells. Therefore, in addition to adoption of the GIDL current read method, which allows a highly sensitive detection for the locally stored charges even with a small amount of charges, we combined separated paths for F-N tunneling during program and erase, which allows improved performance for repeated cycling of program and erase. The F-N tunneling path for programming is performed through the gate-to-drain overlap region, while that for erasing is performed through the gate-to-substrate. Compared with the conventional program/erase operation in which it is performed through the gate and the substrate, the tunnel oxide degradation during the program/erase operation is expected to be significantly suppressed due to the separated path for the programming from the erasing path. Therefore, we expect that the 4-bit TGIF NVM cell has better reliability characteristics related to the endurance and retention with repeated P/E cycling.

To prevent overprogramming and over-erasing, the program and erase performance with various gate and drain voltages is investigated. The disturbance of other bits during program and erase is also checked, while the selected bit is being programmed. The transient characteristics of various program/erase voltage conditions for the 4-bit TGIF NVM cell are shown in Fig. 9. From the simulation result, we expect that the optimized condition for program time $T_P = 1 \times 10^{-6}$ s is $V_G = 5$ V and $V_{DS} = -2$ V, while that for erase time $T_E = 1 \times 10^{-6}$ s is observed to be $V_G = -4$ V and $V_{DS} = 2$ V. In the simulation, Bit2 is selected for program/erase, and Bit1 under the same gate electrode is unselected. It is clear that the vertical electric field is too low to inject the charges to Bit1 in that short time, and, thus, no program/erase disturbance occurs. The selected charge storage node (Bit2), however, can be programmed (and/or erased) quickly at low V_{DS} by the lateral electric field through the gate-to-drain overlap region. Similarly, from the vertical electric field effect simulation result of Bit1, the unselected Bit1 is not programmed (and/or erased) at low V_G and robust to the program/erase disturbance.

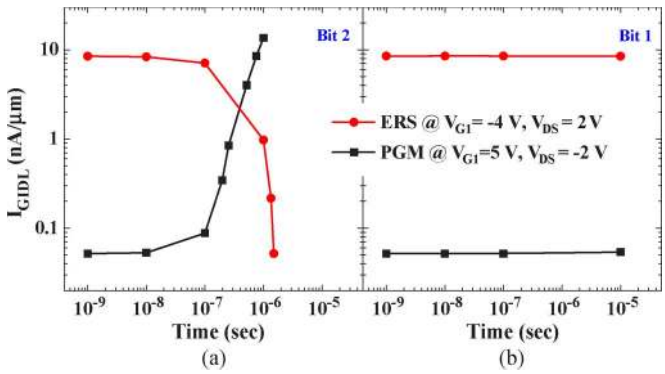

Fig. 9. Time-dependent program/erase characteristics (● = Erase at $V_{G1} = -4$ V and $V_{DS} = 2$ V, ■ = Program at $V_{G1} = 5$ V, $V_{DS} = -2$ V) when (a) Bit2 is selected and (b) Bit1 is unselected.

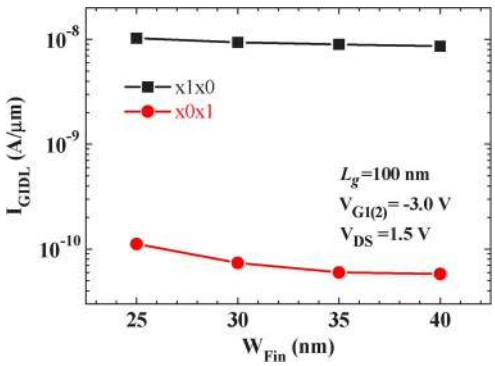

Fig. 10. GIDL current as a function of the fin width for two different states (margin, ● = "x0x1" and ■ = "x1x0") with the smallest GIDL current separation for comparison of the second-bit effect.

IV. The Second-Bit Effects and Reliability of the 4-Bit TGIF NVM Cell

A. Gate Length Scalability and Second-Bit Effects

The programmed charge is sensed by the GIDL current from the common drain electrode in the 4-bit TGIF NVM structure. The bit is sandwiched between the same drain electrodes, while the gate electrode on the opposite side of the selected bit may affect the GIDL current when we sense the logic state of the programmed bit. In order to verify the robustness to the second-bit effect in the 4-bit TGIF SONOS NVM cell structure, additional simulations are performed for various combinations of the fin width. In Fig. 10, the GIDL current is shown as a function of the fin width for the two states with the smallest GIDL current separation (i.e., states "x1x0" and "x0x1"). From the simulation result, it is clear that the second-bit effect by the narrow fin width is negligible, and the separation of the GIDL current is still large enough ($\sim 10^2$ A/μm), although the fin width is scaled down to 25 nm.

In the conventional V_T read method, on the other hand, the second-bit effect becomes severe when the source and the drain get close together with the scaling down of the cell because the programmed charges near the drain affect the substrate energy barrier near the source. The GIDL current method also can be affected by the programmed charges near the source. In order to investigate the second-bit effect in short-channel devices, TCAD simulation was performed for a 4-bit TGIF NVM structure for a short gate length $L_g = 50$ nm. The width of the fin body is scaled down to a proper width to satisfy $L_g > \sim 1.5 \times W_{\rm Fin}$ for $L_g = 50$ nm in order to suppress the degradation of V_T due to the increased SCE [24]. As shown in Fig. 11(b), a considerable separation in the GIDL current is still achieved between the programmed and erased states for $L_g = 50$ nm and $W_{\rm Fin} = 20$ nm. From the result, we expect that the 4-bit TGIF NVM with $L_g = 50$ nm operates without appreciable disturbance. However, three different threshold voltages are observed in the device with $L_g = 50$ nm, while there is negligible V_T separation in the device with $L_g = 100$ nm. This result informs that the conventional V_T read method is significantly subject to the second-bit effect and, therefore, inappropriate to read the charge state in the proposed TGIF NVM cells.

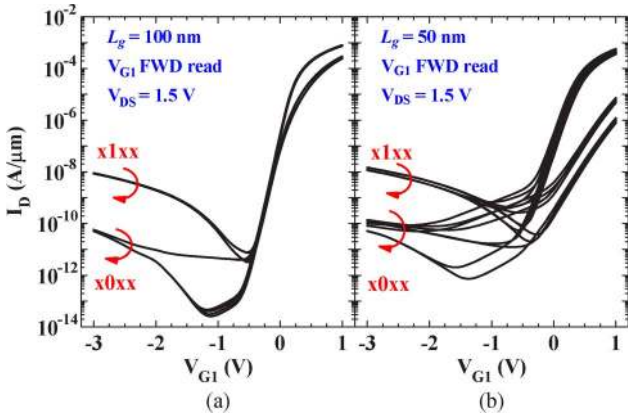

Fig. 11. Comparison of the gate length scalability ($L_g = 100$ nm and 50 nm) at $V_{G1} = -3.0$ V, $V_{DS} = 1.5$ V. (a) I_D–V_{G1} characteristics for $L_g = 100$ nm and $W_{\rm Fin} = 40$ nm. (b) I_D–V_{G1} characteristics for $L_g = 50$ nm and $W_{\rm fin} = 20$ nm.

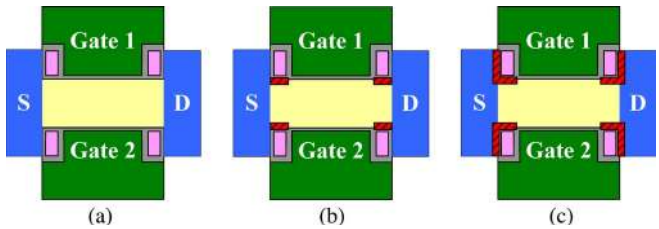

Fig. 12. Schematic of three different interface states for simulation. (a) "Initial states" without any interface state. (b) "Conventional P/E scheme": the interface state ($N_{\rm it} = 1 \times 10^{13}$ cm^{-2}) exists only in the channel regions (red zones) under the ONO gate dielectric stacks. (c) "TGIF P/E scheme": the interface state is set at $N_{\rm it} = 5 \times 10^{12}$ cm^{-2} in both the channel region under the ONO gate dielectric stack and the gate-to-source/drain overlap region next to the ONO gate dielectric stack.

B. Reliability of the 4-Bit TGIF NVM Cells

For comparative investigation of the reliability characteristics of the proposed TGIF NVMs on the P/E schemes, we set three different cases: 1) the initial state as a reference; 2) the conventional P/E scheme; and 3) the "TGIF P/E scheme" with separated P/E paths adopted for enhanced reliability, as shown in Fig. 12. The "initial state," as shown in Fig. 12(a), is a reference set to have no interface state implying no degradation by stresses due to the program and erase processes. The "conventional P/E scheme," as shown in Fig. 12(b), is set to emulate

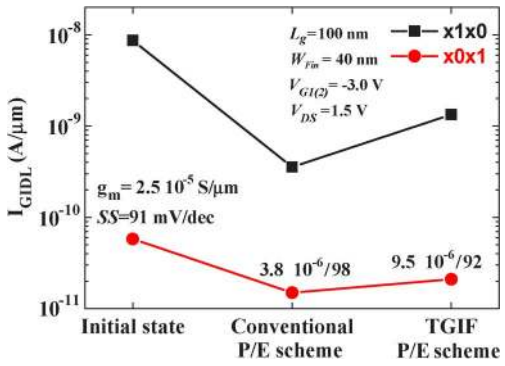

Fig. 13. Comparison of GIDL currents for the interface state effect on the sensing margin (●="x0x1" and ■="x1x0"). "Initial state" has no interface state, and the "conventional P/E scheme" has $N_{\rm it} = 1 \times 10^{13}$ cm^{-2} interface traps only in the channel region. For the "TGIF P/E scheme," the same interface state ($N_{\rm it} = 5 \times 10^{12}$ cm^{-2}) both in the channel region and in the gate-to-source/drain overlap region.

the degraded interface by a conventional program/erase scheme on the proposed TGIF NVMs, i.e., using the same vertical path for F-N tunneling during program and erase processes. The "TGIF P/E scheme," as shown in Fig. 12(c), is set to emulate the degradation of the TGIF NVM cells by using a different F-N tunneling path for program from that for erase. In the "conventional P/E scheme," the interface state exists only in the channel region with $N_{\rm it} = 1 \times 10^{13}$ cm^{-2} [31]. For the "TGIF P/E scheme," the interface state is set at $N_{\rm it} = 5 \times 10^{12}$ cm^{-2} in both the channel region and the gate-to-source/drain overlap region. Here, we simply assumed that the same amount of interface traps between the silicon channel and silicon-oxide (Si-SiO_2) are generated proportionally by the number of the program/erase cyclings.

The device performance parameters of the proposed TGIF NVMs after PE cycling for the "conventional P/E scheme" and the "TGIF P/E scheme" are compared on the transconductance g_m and the subthreshold slope SS. As expected, the interface traps between the channel and the tunnel oxide influence on the direct-current characteristics more than the interface traps between the raised source/drain overlap regions. The interface state in the channel region is expected to influence on the sensing margin, as shown in Fig. 13. Although the "TGIF P/E scheme" has a lower GIDL current than the "initial state," the separation is high enough ($\sim 10^2$ A/μm) to identify the logic state of each bit. However, the "conventional P/E scheme" shows a small sensing window as well as small separation of the GIDL current. From these results, the proposed program/erase scheme with different paths for program and erase in the 4-bit TGIF NVM cells is expected to be very effective to improve the reliability characteristics. We also note that the advantage of the TGIF P/E scheme in the simulation result would be underestimated since the interface trap density for the "TGIF P/E scheme" is set at half of that for the "conventional P/E scheme," although it is generally known that the generation density of the interface states follows a power law for the stress time or the total number of P/E cyclings [32]. In other words, for the 4-bit TGIF NVM cells, the "TGIF P/E scheme" in practical implementation guarantees much better reliability compared with that of the "conventional P/E scheme."

V. Summary and Conclusion

We have proposed a novel 4-bit T-gate and I-shape FinFET SONOS-type NVM for high storage density with better reliability. In order to improve the immunity to the second-bit effect, the GIDL current method has been employed for sensitive detection of the stored bits at low voltage. Separated paths for program and erase with a modified F-N tunneling mechanism have been adopted for enhanced reliability characteristics after repeated P/E cycling. The operation and the performance have been investigated by using a 2-D TCAD simulation with a crossed cell array structure and a common word line for high storage density. Although the operation and electrical characteristics have been confirmed by TCAD simulation tools, we have fully considered a process flow for practical implementation. In order to simplify the process and improve a possible mismatch caused by misalignment during fabrication, we have adopted a self-alignment process for the four physically separated charge storage nodes.

Utilizing the BTBT current, we have employed the GIDL current method for enhanced charge sensing of the logic state in each bit. The GIDL current read method has been observed to be very sensitive to locally stored charges, and, thus, a small amount of charges is enough to determine the charge state. We have obtained a large sensing window from the 4-bit TGIF NVM for the gate length $L = 50$ nm. Therefore, the program/erase operation can be performed at low voltage for a short period of time due to the high sensitivity of the GIDL current method. It is also immune to the second-bit effect and allows an aggressive scaling down of the proposed 4-bit TGIF NVM cell for high-density memory applications. In order to improve the reliability characteristics, we have used different paths for program and erase operation. A modified F-N tunneling mechanism has been used for programming through the gate-to-source/drain overlap regions (along the channel length direction) and the BBHHI for erasing through the gate-to-substrate regions (along the channel width direction). Through 2-D TCAD simulation, we have confirmed that the TGIF NVM cell with separated paths for program and erase has better reliability characteristics in the sensing margin, transconductance, and subthreshold slope. Combining the modified F-N tunneling mechanism with different paths for the P/E and GIDL current read method, the proposed 4-bit TGIF NVM device structure is expected to be very useful for high-density storage applications in the next generation.

References

[1] R. Degraeve, F. Schuler, M. Lorenzini, D. Wellekens, P. Hendrickx, J. Van Houdt, L. Haspeslagh, G. Groeseneken, and G. Tempel, "Analytical model for failure rate prediction due to anomalous charge loss of Flash memories," in *IEDM Tech. Dig.*, 2001, pp. 32.1.1–32.1.4.

[2] M. L. French, C.-Y. Chen, H. Sathianathan, and M. H. White, "Design and scaling of a SONOS multi-dielectric device for nonvolatile memory applications," *IEEE Trans. Compon., Packag., Manuf. Technol. A*, vol. 17, no. 3, pp. 390–397, Sep. 1994.

[3] 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.

[4] 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.

[5] J.-D. Lee, S.-H. Hur, and J.-D. Choi, "Effects of floating-gate interference on NAND Flash memory cell operation," *IEEE Electron Device Lett.*, vol. 23, no. 5, pp. 264–266, May 2002.
[6] K. Kim, "Technology for sub-50 nm DRAM and NAND Flash manufacturing," in *IEDM Tech. Dig.*, 2005, pp. 323–326.
[7] B. Eitan, P. Pavan, I. Bloom, E. Aloni, A. Frommer, and D. Finzi, "NROM: A novel localized trapping, 2-bit nonvolatile memory cell," *IEEE Electron Device Lett.*, vol. 21, no. 11, pp. 543–545, Nov. 2000.
[8] C. C. Yeh, W. J. Tsai, T. C. Lu, S. K. Cho, T. Wang, S. Pan, and C. Y. Lu, "A modified read scheme to improve read disturb and second-bit effect in a scaled MXVAND Flash memory cell," in *Proc. NVSM Workshop*, 2003, pp. 44–45.
[9] N. Lindert, L. Chang, Y.-K. Choi, E. H. Anderson, W.-C. Lee, T.-J. King, J. Bokor, and C. Hu, "Sub-60-nm quasi-planar FinFETs fabricated using a simplified process," *IEEE Electron Device Lett.*, vol. 22, no. 10, pp. 487–489, Oct. 2001.
[10] P. Xuan, M. She, B. Harteneck, A. Liddle, J. Bokor, and T.-J. King, "FinFET SONOS Flash memory for embedded applications," in *IEDM Tech. Dig.*, 2003, pp. 609–612.
[11] M. Specht, M. Specht, U. Dorda, L. Dreeskomfeld, J. Kretz, F. Hofmann, M. Stadele, R. J. Luyken, W. Rosner, H. Reisinger, E. Landgraf, T. Schulz, J. Hartwich, R. Kommling, and L. Risch, "20 nm tri-gate SONOS memory cells with multi-level operation," in *IEDM Tech. Dig.*, 2004, pp. 1083–1084.
[12] J.-R. Hwang, T.-L. Lee, H.-C. Ma, T.-C. Lee, T.-H. Chung, C.-Y. Chang, S.-D. Liu, B.-C. Perng, J.-W. Hsu, M.-Y. Lee, C.-Y. Ting, C.-C. Huang, J.-H. Wang, J.-H. Shieh, and F.-L. Yang, "20 nm gate bulk-FinFET SONOS Flash," in *IEDM Tech. Dig.*, 2005, pp. 154–157.
[13] M. Specht, R. Kömmling, F. Hofmann, V. Klandzievski, L. Dreeskornfeld, W. Weber, J. Kretz, E. Landgraf, T. Schulz, J. Hartwich, W. Rösner, M. Städele, R. J. Luyken, H. Reisinger, A. Graham, E. Hartmann, and L. Risch, "Novel dual bit tri-gate charge trapping memory devices," *IEEE Electron Device Lett.*, vol. 25, no. 12, pp. 810–812, Dec. 2004.
[14] M. Fukuda, T. Nakanishi, and Y. Nara, "New nonvolatile memory with charge-trapping sidewall," *IEEE Electron Device Lett.*, vol. 24, no. 7, pp. 490–492, Jul. 2003.
[15] B. Y. Choi, B. Y. Choi, B.-G. Park, J. D. Lee, H. Shin, Y. K. Lee, K. H. Bai, D.-D. Kim, D.-W. Kim, C.-H. Lee, and D. Park, "Reliable 2-bit/cell NVM technology using twin SONOS memory transistor," *Electron. Lett.*, vol. 41, no. 19, pp. 1086–1087, Sep. 2005.
[16] C.-E. Juang, H.-M. Chen, H.-C. Lai, Y.-J. Chen, Y.-C. King, and C. J. Lin, "A new self-aligned nitride MTP cell with 45 nm CMOS fully compatible process," in *IEDM Tech. Dig.*, 2007, pp. 91–94.
[17] K.-T. Chang, W.-M. Chen, C. Swift, J. M. Higman, W. M. Paulson, and K.-M. Chang, "A new SONOS memory using source-side injection for programming," *IEEE Electron Device Lett.*, vol. 19, no. 7, pp. 253–255, Jul. 1998.
[18] Y. Hayashi, S. Ogura, T. Saito, and T. Ogura, "Twin MONOS cell with dual control gates," in *VLSI Symp. Tech. Dig.*, 2000, pp. 122–123.
[19] B. Eitan, G. Cohen, A. Shappir, E. Lusky, A. Givant, M. Janai, I. Bloom, Y. Polansky, O. Dadashev, A. Lavan, R. Sahar, and E. Maayan, "4-bit per cell NROM reliability," in *IEDM Tech. Dig.*, 2005, pp. 539–542.
[20] M. Janai, "Data retention, endurance and acceleration factors of NROM devices," in *Proc. IRPS*, 2003, pp. 502–505.
[21] M. Janai, B. Eitan, A. Shappir, E. Lusky, I. Bloom, and G. Cohen, "Data retention reliability model of NROM nonvolatile memory products," *IEEE Trans. Device Mater. Rel.*, vol. 4, no. 3, pp. 404–415, Sep. 2004.
[22] A. Padilla, K. Shin, T.-J. K. Liu, J. W. Hyun, I. Yoo, and Y. Park, "Dual-bit gate-sidewall storage FinFET NVM and new method of charge detection," *IEEE Electron Device Lett.*, vol. 28, no. 6, pp. 502–505, Jun. 2007.
[23] A. Padilla, S. Lee, D. Carlton, and T.-J. K. Liu, "Enhanced endurance of dual-bit SONOS NVM cells using the GIDL read method," in *VLSI Symp. Tech. Dig.*, 2008, pp. 142–143.
[24] I. De and C. M. Osburn, "Impact of super-steep-retrograde channel doping profiles on the performance of scaled devices," *IEEE Trans. Electron Devices*, vol. 46, no. 8, pp. 1711–1713, Aug. 1999.
[25] L. Chang, Y.-K. Choi, D. Ha, P. Ranade, S. Xiong, J. Bokor, C. Hu, and T.-J. King, "Extremely scaled silicon nano-CMOS devices," *Proc. IEEE*, vol. 91, no. 11, pp. 1860–1873, Nov. 2003.
[26] Synopsys, Inc., Mountainview, CA, v. 2006.12 Sentaurus User's Manual.
[27] Z. Liu, C. Lee, V. Narayanan, G. Pei, and E. C. Kan, "Metal nanocrystal memories, Part I: Device design and fabrication," *IEEE Trans. Electron Devices*, vol. 49, no. 9, pp. 1606–1613, Sep. 2002.
[28] Z. Liu, C. Lee, V. Narayanan, G. Pei, and E. C. Kan, "Metal nanocrystal memories, Part II: Electrical characteristics," *IEEE Trans. Electron Devices*, vol. 49, no. 9, pp. 1614–1622, Sep. 2002.
[29] C. Wang, M. O. Ahmad, and M. N. S. Swamy, "A CMOS current amplifier for very weak current operations," in *Proc. ESSCIRC*, 2002, pp. 751–754.
[30] A. K. Sharma, *Advanced Semiconductor Memories*. New York: Wiley-Interscience, 2003, pp. 349–350.
[31] A. Furnemont, M. Rosmeulen, K. van der Zanden, J. Van Houdt, K. De Meyer, and H. Maes, "Root cause of charge loss in a nitride-based localized trapping memory cell," *IEEE Trans. Electron Devices*, vol. 54, no. 6, pp. 1351–1359, Jun. 2007.
[32] S. Mahapatra, M. A. Alam, P. Bharath Kumar, T. R. Dalei, D. Varghese, and D. Saha, "Negative bias temperature instability in CMOS devices," *Microelectron. Eng.*, vol. 80, pp. 114–1121, Jun. 2005.

Sunyeong Lee (S'07) received the B.S. and M.S. degrees in electrical engineering from Kookmin University, Seoul, Korea, in 2007 and 2010, respectively.

From September 2007 to August 2008, he was with the Department of Electrical Engineering and Computer Science, University of California, Berkeley, where he was involved in the characterization of multibit nitride-based charge-trapping Flash memory, as well as the design and development of 3-D nanoscale FETs. He is currently with Kookmin University. His current research interests include design, characterization, and modeling of Si nanoelectronic devices, Flash memory devices, DRAM devices, and high-performance devices and integrated circuits.

Yong Woo Jeon received the B.S. degree in electrical engineering in 2009 from Kookmin University, Seoul, Korea, where he is currently working toward the M.S. degree in the Department of Electrical Engineering, under the supervision of Prof. D. M. Kim and Prof. D. H. Kim.

His current research mainly focuses on amorphous oxide semiconductor TFTs and their device physics.

Tsu-Jae King Liu (SM'00–F'07) received the B.S., M.S., and Ph.D. degrees in electrical engineering from Stanford University, Stanford, CA, in 1984, 1986, and 1994, respectively.

In 1992, she joined the Xerox Palo Alto Research Center, Palo Alto, CA, as a member of the Research Staff to research and develop polycrystalline-silicon thin-film transistor technologies for high-performance flat-panel displays. In August 1996, she joined the faculty of the University of California, Berkeley, where she is currently the Conexant Systems Distinguished Professor of electrical engineering and computer sciences and the Associate Dean for Research in the College of Engineering. Her research activities are currently in nanometer-scale integrated-circuit devices and technology, as well as materials, processes, and devices for integrated microsystems.

Dr. Liu was an Editor for the IEEE Electron Devices Letters from 1999 to 2004. She has served on committees for many technical conferences, including the International Electron Devices Meeting and the Symposium on VLSI Technology. She was a recipient of several awards, including the DARPA Significant Technical Achievement Award (2000) for the development of the FinFET and the IEEE Kiyo Tomiyasu Award (2010) for contributions to nanoscale MOS transistors, memory devices, and MEMS devices. She has served on committees for many technical conferences, including the International Electron Devices Meeting and the Symposium on VLSI Technology. She was an Editor for the IEEE Electron Devices Letters from 1999 to 2004.

Dae Hwan Kim (M'08) received the B.S., M.S., and Ph.D. degrees in electrical engineering from Seoul National University, Seoul, Korea, in 1996, 1998, and 2002, respectively.

From 2002 to 2005, he was with Samsung Electronics Company, Ltd., Kyungki-Do, Korea, where he contributed to the design and development of 92-nm DDR DRAMs and 80-nm DDR2 DRAMs. In 2005, he joined the School of Electrical Engineering, Kookmin University, Seoul, as an Assistant Professor. His current research interests include nanoscale CMOS devices and circuits, future memory devices, low-voltage and low-power nano-IC, and Si quantum devices and their applications.

Dong Myong Kim (S'86–M'88) received the B.S. (*magna cum laude*) and M.S. degrees in electronics engineering from Seoul National University, Seoul, Korea, in 1986 and 1988, respectively, and the Ph.D. degree in electrical engineering from the University of Minnesota, Twin Cities, MN, in 1993.

From February 1988 to August 1989, he was with the Division of Electronic Engineering, Korea Institute of Science and Technology, Seoul, where he was involved in the characterization and modeling of microwave devices and integrated circuits. Since 1993, he has been with the School of Electrical Engineering, Kookmin University, Seoul. His current research interests include fabrication, characterization, and modeling of nanostructure silicon devices, thin-film transistors, biosensors, III–V compound semiconductor devices, volatile and nonvolatile memories, and CMOS RF circuits.