A Novel Self-Aligned Highly Reliable Sidewall Split-Gate Flash Memory

Caleb Yu-Sheng Cho, *Member, IEEE*, Ming-Jer Chen, *Senior Member, IEEE*, Chiou-Feng Chen, Prateep Tuntasood, Der-Tsyr Fan, and Tseng-Yi Liu

***Abstract*—A self-aligned sidewall split-gate Flash memory cell is fabricated with overerase immunity. Particularly, the sidewall corner of the floating-gate is deliberately rounded to release the electric field lines encountered in the poly-to-poly erase. The unit cell size of $12.7\ F^2$ (F is the feature size), formed in a 32-Mb NOR architecture, and the acceptable erase speed of 20 ms for block erase (512 K bits, 16 pages) are quite competitive. Endurance cycles up to 10^5 confirm the novel cell to be highly reliable as compared with the conventional source-side erase scheme. The bake experiment at 250 °C before and after program/erase cycles indicates the cell not only free of extrinsic defects in the manufacturing process but also experiencing excellent retention characteristics. Disturb effects during the programming and read-out operations are examined in detail and the operating conditions for disturbs inhibition are readily determined. We eventually elaborate on the differences between the proposed cell structure and existing ones, as well as on the NAND architecture application.**

***Index Terms*—Flash memory, MOSFETs, NAND, NOR, overerase, poly erase, sidewall, source-side injection, split-gate.**

I. Introduction

SPLIT-GATE Flash memories with a sidewall poly 3 gate [1], [2] have recently received significant attention due to promising advantages of low power consumption, high injection efficiency, less susceptibility to short-channel effects (drain turn-on and punch-through), and overerase immunity. Here, the overerase immunity is due to the built-in select-gate transistor, which can effectively get rid of the on-chip erase procedures traditionally used to resolve the overerase problems in stacked gate cell. On the other hand, the unit cell area is considered to be competitive ($\sim 14\ F^2$ [2]; F is the feature size) relative to that of stacked gate cell (typical $\sim 10\ F^2$).

In this paper, the sidewall corner of the floating-gate (FG) in a sidewall split-gate Flash memory cell is deliberately rounded [3], as experimentally carried out in a 32-Mb NOR Flash memory with a unit cell size of $12.7\ F^2$. The rounded shape can effectively release the field lines at the FG corner during poly-to-poly erase and as a result, the damage to interpoly oxide is minimized. Such improvement leads to excellent endurance characteristics, as verified in comparison with the conventional source-side erase scheme. At the same time, the acceptable erase speed is achieved. The retention bake test is addressed before and after P/E cycling. Disturb effects are examined in detail and the operating conditions for disturbs inhibition are determined. Also elaborated are the differences relative to existing source-side injection and/or split-gate structures, as well as the potential application in the NAND architecture.

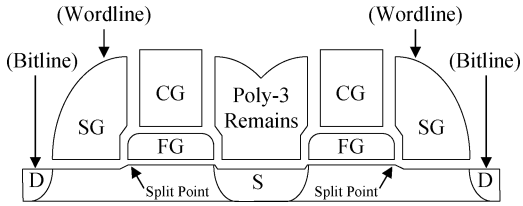

Fig. 1. Schematic cross-sectional view of the novel cell along cell channel direction.

II. Experimental

A self-aligned sidewall split-gate 32-Mb NOR Flash memory is fabricated in 0.15-μm technologies with shallow trench isolation. Some key processes are described as follows. Like the conventional procedures in stacked-gate device, the floating gate (poly 1 layer) in this paper is self-aligned and then etched by using the control gate (CG; poly 2 layer) as a hard mask. The CG and subsequent deposited dielectric film together serve as another mask to facilitate the formation of a self-aligned select gate (SG) while undergoing anisotropic dry etching on poly 3 layer. To enhance the potential coupling toward the FG during the programming, the remaining poly 3 over the source area is metal-connected to the underlying source diffusion. Actually, this kind of contact connection is implemented over several cells along wordline (W/L) for suppression of the area overhead. Fig. 1 schematically shows the cross section of this novel cell along the cell channel direction. Here, the "split point" divides the cell into a stacked-gate cell and a select-gate transistor; and the W/L and bitline (B/L) are defined by the select gate (SG) and drain (D), respectively. The overerase immunity is realized with the built-in select-gate transistor. The floating-gate sidewall corner next to the interpoly oxide is deliberately rounded. Fig. 2 shows the corresponding transmission electron microscopy (TEM) image along the cell channel direction. Some of the important cell dimensional parameters

Manuscript received April 19, 2005; revised December 2, 2005. This work was supported by the National Science Council, Taiwan, R.O.C., under Contract NSC 90-2218-E-009-043. The review of this paper was arranged by Editor R. Shrivastava.

C. Y.-S. Cho, D.-T. Fan, and T.-Y. Liu are with Actrans System Inc., Hsinchu 300, Taiwan, R.O.C.

M.-J. Chen is with the Department of Electronics Engineering, National Chiao-Tung University, Hsinchu 300, Taiwan, R.O.C. (e-mail: chenmj@faculty.nctu.edu.tw).

C.-F. Chen and P. Tuntasood are with Actrans System Inc., Santa Clara, CA 95054 USA.

Digital Object Identifier 10.1109/TED.2005.863764

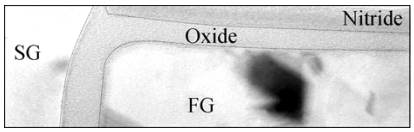

Fig. 2. Localized cross-sectional TEM image showing the rounded corner of FG.

TABLE I
KEY CELL DIMENSIONAL PARAMETERS

Unit Cell Size	0.28 μm^2
CG over FG along Channel	0.15 μm
Channel Width	0.18 μm
Floating-Gate Oxide Thickness	9.5 nm

are listed in Table I. The additional select gate for the purpose of overerase immunity increases the unit cell size to 12.7 F^2, but still being competitive to that (e.g., 10 F^2) of stacked gate cell.

The array size is divided into page and block with 32 K cells and 16 pages, respectively. In the selected page, CG and source are in common potential throughout the whole page; and cells can be either selected or deselected by both W/L and B/L. Fig. 3(a) depicts the page configuration of the array architecture, which is arranged in the row organization with common B/Ls shared by all pages. According to the coded address, individual page is selected through both control-gate decoder and source switch; and memory cells in the selected page are selected by activating the W/L and B/L decoders. On the other hand, deselected pages are all biased at *grounded* $V_{\rm SG}$, $V_{\rm CG}$ and V_S, which can greatly alleviate those disturbs from cycled program/erase operations [4]. In addition, since the drain voltage used in this work is small, the traditional overhead penalty due to extra local B/L decoders is eliminated.

III. CELL OPERATIONS

Programming is performed by means of source-side channel-hot-electron injection: applying typical 10 V to the CG, 3.3 V to the source (S), 0 V to the drain (D), and 1.5 V to the select gate (SG). Under these conditions, the floating-gate voltage ($V_{\rm FG}$) is capacitively coupled to a high level enough to strongly turn on a conductive channel beneath, which in turn gives rise to an inversion layer electrically connecting the source voltage (V_S) to the proximity of the split point as depicted in Fig. 1. As mentioned in the Experimental section, on the other hand, the poly 3 over the source area is metal-connected to the source diffusion and thereby the potential coupling is enhanced toward FG during programming. Operating voltages for selected page are given in Fig. 3(b) that displays eight 4-Kb W/Ls (poly 3) and eight CG lines (poly 2). The cell C2 is selected for programming. For other cells C1 and C3 in selected W/L ($V_{\rm SG} = 1.5$ V), program is inhibited by applying 1 V on corresponding B/Ls, and cells in the deselected W/Ls, e.g., C4, C5, and C6, are all inhibited by $V_{\rm SG} = 0$ V. Note that in the deselected pages, $V_{\rm SG}$, $V_{\rm CG}$ and V_S are all grounded under programming. This type of special configuration greatly alleviates propagation of program disturb among pages, indicating that each page can be electrically isolated from other pages.

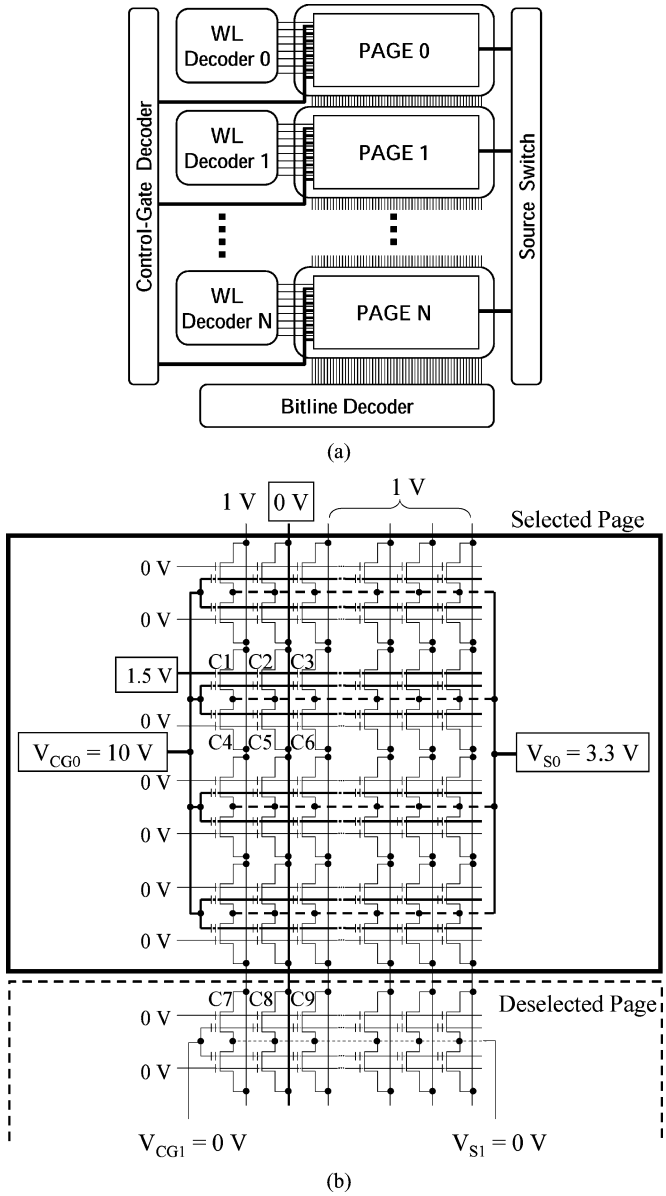

Fig. 3. (a) NOR Flash memory array architecture used in the study. The array is arranged in the row organization with B/Ls commonly shared by all pages. Page is selected or deselected by CG and source switch. Note that not all pages are depicted here. (b) Operating voltages in selected page during programming in NOR architecture. C2 is the selected cell. A single page is made up of 8 W/Ls, 4096 B/Ls, and common CG and source in this page.

Poly-to-poly Fowler–Nordheim (FN) electron tunneling (i.e., 11 V on SG and 0 V on CG) from the floating gate toward the neighboring select gate (W/L) is employed for erasing. The erasing size is quite flexible, ranging from a smallest size of 4 Kbits (one W/L) to the whole chip size (32 Mbits). Memory cells of selected block (512 Kbits, 16 pages) can be over-erased to negative threshold voltage ($V_{\rm TH}$) in 20 ms, without use of extra on-chip erase verify procedures due to built-in overerase immunity. The deselected pages are biased in grounded $V_{\rm SG}$, $V_{\rm CG}$, and V_S, and open $V_{\rm D}$ to ensure erase inhibition.

For an over-erased cell in deselected W/L ($V_{\rm SG} = 0$ V), the undesirable leakage current is eliminated as experimentally evidenced in Fig. 4. Finally, the cell is read out by applying 4 V on SG, 1.5 V on drain, and 0 V on CG. The typical read-out current

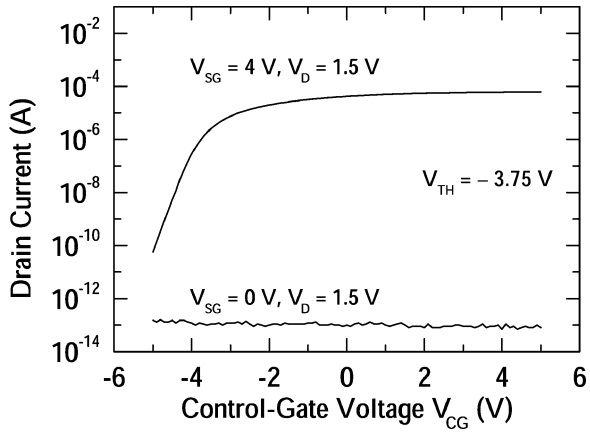

Fig. 4. Read-out characteristics of an over-erased memory cell ($V_{\rm TH} = -3.75$ V) for selected and deselected cases. The deselected case ($V_{\rm SG} = 0$) yields good overerase immunity under read-out operation.

TABLE II
TYPICAL OPERATING VOLTAGES AND PULSE DURATIONS IN SELECTED PAGE. THE "SEL" AND "DESEL" DENOTE SELECTED AND DESELECTED, RESPECTIVELY. NOTE THAT CG AND SOURCE ARE COMMON IN ONE PAGE BUT NOT COMMON FOR THE WHOLE ARRAY

	Time	CG	Source	SG (W/L)		Drain (B/L)	
Program	10 μs	10 V	3.3 V	Sel	DeSel	Sel	DeSel
				1.5 V	0 V	0 V	1 V
Erase	10 *ms*	0 V	0 V	11 V		OPEN	
Read	70 *ns*	0 V	0 V	4 V	0 V	1.5 V	0 V

is $30 \sim 40$ μA for over-erased cells. The pulse voltages and durations used here for selected (denoted as "Sel") and deselected cells (denoted as "DeSel") are listed in Table II.

IV. CHARACTERISTICS OF THE NOVEL CELL

Fig. 5 shows the gate and source current measured versus select-gate voltage $V_{\rm SG}$ in a dummy transistor, which is the memory cell with shorted CG and FG. The maximum gate current reaches at $V_{\rm SG} = 1.3$ V for source voltage $V_S = 3.3$ V. Considering the process variations cell by cell while actually accounting for the dependence of $V_{\rm SG}$ on gate current (that is, a smaller gate current variation to the right of the peak $V_{\rm SG} \approx 1.3$ V in Fig. 5), the $V_{\rm SG}$ of 1.5 V was exploited in this paper. The corresponding source current is merely 10 μA, which is highly favorable for programming of several words.

Fig. 6 shows the measured programming characteristics with CG voltage as a parameter. It can be seen that for the specific $V_{\rm CG} = 10$ V, initially the threshold voltage rapidly increases until entering certain transition region and subsequently the increase rate of threshold voltage is lowered. The physical interpretation is that in the beginning of programming time, FG is highly positively coupled from the source via the extended channel underneath FG; consequently, FG channel is operated in linear region. As more electrons are injected onto FG, $V_{\rm FG}$ becomes more negative and the depletion region is gradually formed near source side; that is, FG channel enters into saturation region of operation [5]. Obviously, to maintain FG channel in linear region during the programming period, the target programmed $V_{\rm TH}$ is appropriately set at 2 V for $V_{\rm CG} = 10$ V. The

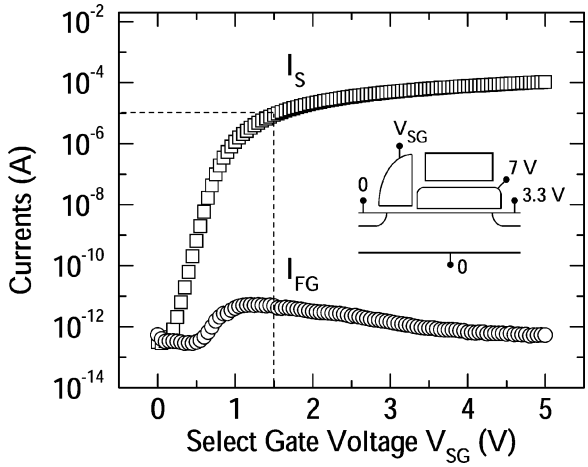

Fig. 5. Gate ($I_{\rm FG}$) and source (I_S) current measured versus select-gate voltage in a dummy transistor. For programming, $V_{\rm SG} = 1.5$ V is chosen by accounting for the process variations cell by cell as well as the $V_{\rm SG}$-dependence of $I_{\rm FG}$ on $V_{\rm SG}$.

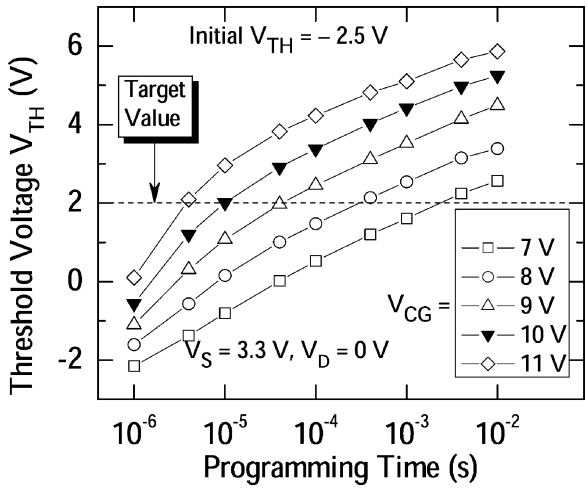

Fig. 6. Programming characteristics of the novel cell in the 0.15-μm technologies. To meet the target value, $V_{\rm CG} = 10$ V, $V_S = 3.3$ V, and $V_D = 0$ V with 10-μs duration are chosen throughout this paper.

corresponding programming time is 10 μs, which appears to fall within the threshold voltage rapidly increasing region.

Fig. 7 shows the measured erase characteristics for various select-gate voltages. Although the rounded FG sidewall considerably lowers the field strength at FG corner during poly-to-poly erase, the erase characteristics of the cell in Fig. 7 reveal that an erase time of 10 ms for the target value of -2.5 V is quite acceptable. To further cover cell erase behaviors in block or even the whole chip, a longer time should be expected to avoid possible erase failure. Erase speed of 20 ms for a block size (512 Kb, 16 pages) in this paper is still competitive.

V. RELIABILITY ISSUES

Reliability issues such as endurance characteristics, retention bake, and disturb effects are addressed in this section. The endurance characteristics focus on the threshold voltage shift with increasing program/erase (P/E) cycles. Retention bake experiment is to record the FG-charge variation occurring in the absence of external voltages. The disturb effects encountered in

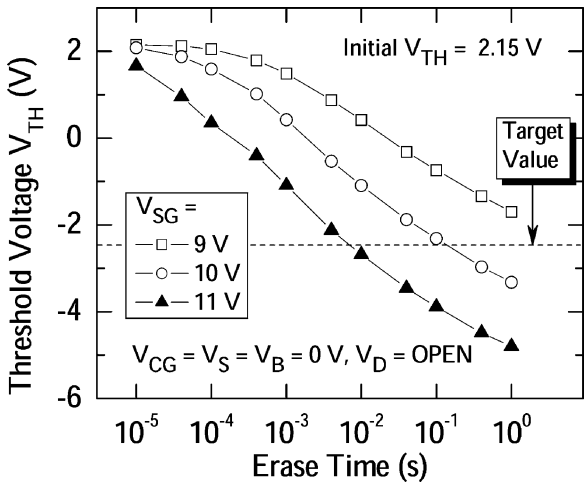

Fig. 7. Erase characteristics of the novel cell in the 0.15-μm technologies. To meet target value, $V_{\rm SG}$ = 11 V with 10 ms duration are chosen in poly-to-poly erase operation.

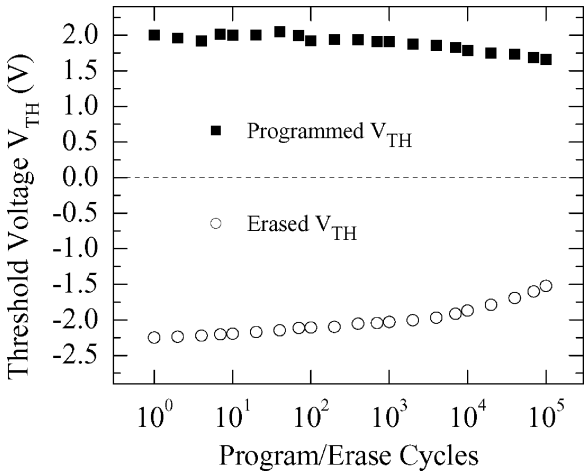

Fig. 8. Endurance characteristics measured on the novel cell. The logic discrimination between two states can be seen.

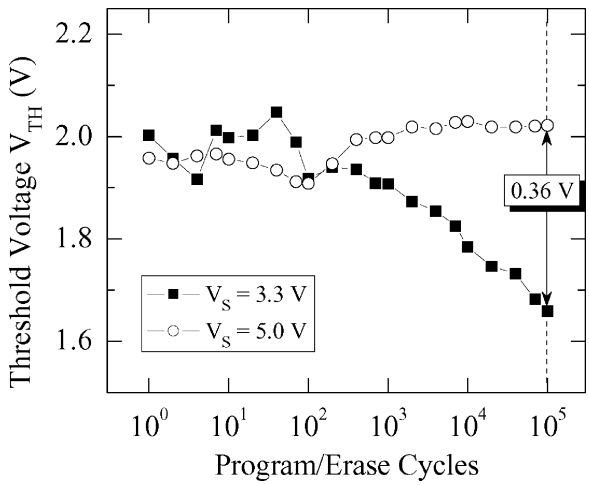

Fig. 9. Endurance characteristics of high $V_{\rm TH}$ for two source voltages of 3.3 and 5 V.

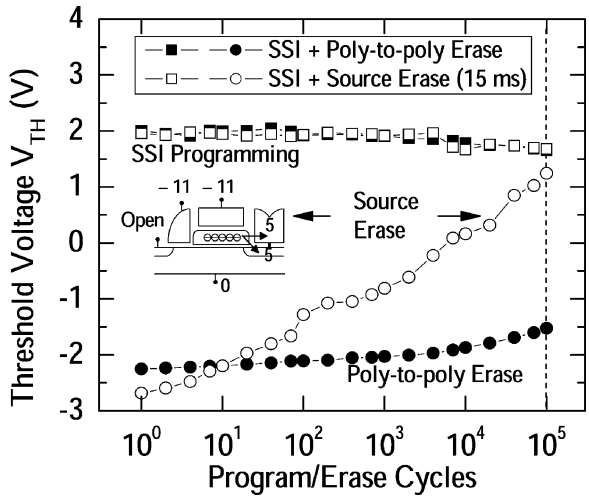

Fig. 10. Endurance characteristics of the novel cell in two erase schemes. The source-side injection programming is applied to both cases. The inset shows the bias configuration for source-side erase.

the program and read-out operations are examined extensively, from which the suitable voltages are determined.

A. Endurance Characteristics

Fig. 8 shows the measured endurance characteristics in terms of high threshold level and low threshold level until P/E cycles of 10^5. It can be seen that (i) the downward shift is 0.35 V for the high threshold state; and (ii) the upward shift in the low threshold state is strikingly small, i.e., 0.7 V. Such small threshold closure in low state manifests the potential of the rounded FG sidewall. The threshold shifts in Fig. 8 can be explained by the mechanism of trapped electrons in cell gate oxide near hot-electron injection point (i.e., the split point in Fig. 1) during programming and in interpoly oxide during poly-to-poly erasing [6]. It is also noteworthy that owing to the weak dependence of channel current on $V_{\rm FG}$ for source-side injection device [5], the low $V_{\rm TH}$ closure does not affect the logic discrimination of sense amplification too much.

1) Endurance Behavior at V_S = 3.3 and 5 V: The endurance characteristics at a higher V_S of 5 V in source-side injection (SSI) show different behavior from that at $V_S = 3.3$ V, as demonstrated in Fig. 9. This figure is obtained by adjusting the duration of 5-V program pulse in such a way to produce comparable threshold voltages in the initial P/E cycles. Indeed, Fig. 9 reveals anomalous characteristics not reported elsewhere: for P/E cycles starting from about 2×10^2, the high threshold voltage state for V_S = 3.3 V dramatically decreases with increasing P/E cycles whereas it relatively remains constant for higher V_S of 5 V. One plausible explanation, which can be found in [6], for the former is that electron trapping in the oxide near the injection point increases with increasing P/E cycle, which in turn lowers the amount of hot electron injection onto the floating gate. The decline rate of hot electron injection exceeds the accumulation rate of electron trapping, resulting in a net decrease in threshold voltage with increasing P/E cycles. The similar argument essentially can apply to the data of $V_S = 5$ V in Fig. 9 but in this case the decline rate of hot electron injection is almost equal to the accumulation rate of electron trapping.

2) Source-Side Erase Versus Poly-to-Poly Erase: Fig. 10 shows the measured endurance characteristics from the poly-to-poly erase and source-side erase. The source-side

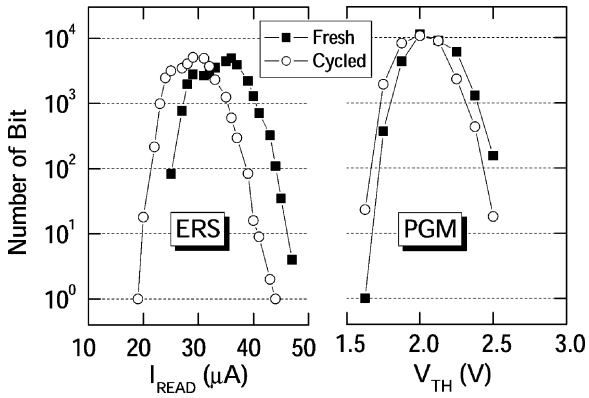

Fig. 11. Distributions of over-erased and programmed cells in terms of $I_{\rm READ}$ and $V_{\rm TH}$ before and after 3×10^4 P/E cycles in 32-Kb arrays, respectively.

injection programming is used for both cases. As expected, the source-side erase scheme results in a worse closure with increasing P/E cycles than poly-to-poly erase. The origin of such large difference is that the source junction is not particularly graded to accommodate source-side erase, as originally cited in [7]. There are two principal reasons for favoring poly-to-poly erase in this work. The first reason is the smaller unit cell size. Because the source implant in poly-to-poly erase is not doped for source-side erase, a smaller unit cell size (and thereby the higher cell densities) is expected. The second reason is that the number of on-chip charge pumps as adopted in source-side erase case is more than that in poly-to-poly erase. Due to negative CG bias used in source-side erase, two positive charge pumps (CG and S) and one negative charge pump (CG) are needed during programming and erasing, respectively. However, only two positive charge pumps are demanded in poly-to-poly erase, making possible realization of smaller die size. Therefore, the poly-to-poly erase method is exploited through this paper.

Fig. 11 displays the distributions of threshold voltage $V_{\rm TH}$ and read-out current $I_{\rm READ}$ respectively from programmed and erased cells in a 32-Kb array (one page) before and after 3×10^4 P/E cycles. It is noteworthy that the whole page programming is accomplished by 10-μs programming byte by byte, but with a 20-ms erasure in the whole page. The longer erase pulse duration is needed due to consideration of process variations. Two interesting statistical properties concerning the read-out current can be drawn. First, after P/E cycles the distribution shape of read current remains intact, except a slight shift of about 5 μA toward the decreasing read current direction. Second, the smallest read-out current after P/E cycles is around 20 μA, which is still large enough for logic discrimination. Also noticed in Fig. 11 is the excellent statistical distribution of post-P/E threshold voltage relative to that of fresh devices. Again, Fig. 11 corroborates the fabricated cells to be statistically highly reliable.

B. Retention Bake

First of all, a 24-h retention bake at 250 °C [8], [9] is performed on the programmed cells in 4-Mb array and the erased cells in 8-Mb array. The resulting retention characteristics are given in Fig. 12(a). This figure exhibits not only a small shift

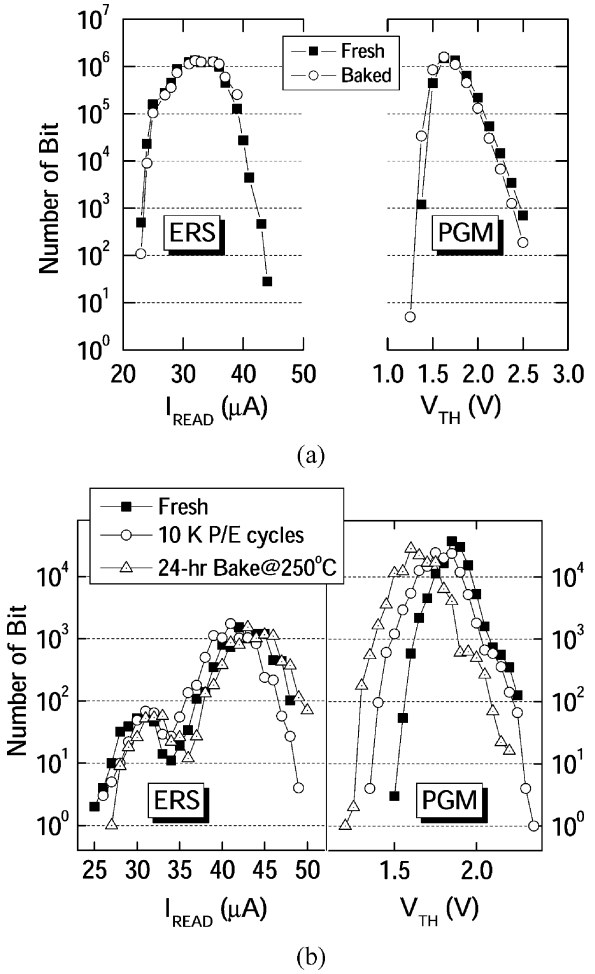

Fig. 12. (a) Retention characteristics before and after 250 °C, 24-hour bake for over-erased and programmed cells in 4- and 8-Mb arrays in terms of $I_{\rm READ}$ and $V_{\rm TH}$, respectively; and (b) retention characteristics for ERS and PGM cells in 8- and 32-Kb arrays, respectively, created from 1) fresh case, 2) post-10^4 P/E cycling cells with no bake, and 3) post-10^4 P/E cycling cells undergoing 24-hr, 250 °C bake.

in both $V_{\rm TH}$ and $I_{\rm READ}$ but also the insignificant change in the distribution after the retention bake. Therefore, it is claimed that the underlying device is free of extrinsic defects or equivalently the oxide thinning in the manufacturing process [10]. The reasons are that if the oxide thinning were not absent, then the intrinsic oxide traps created in the programmed or erased cells would sum up to form a localized critical leakage path in the oxide part having extrinsic defects, which would in turn produce a noticeable change in retention distributions before and after programming or erasing. Furthermore, another 24-h, 250 °C bake experiment is carried out on programmed cells in 32-Kb array and erased cells in 8-Kb array both undergoing 10^4 P/E cycling. The resulting retention bake statistical distributions are plotted in Fig. 12(b). Also given in this figure are those with no bake for comparison. Evidently, it can be drawn that 1) the post-P/E cycling curve with no bake is slightly shifted relative to the fresh array; and 2) again a slight shift appears in post-cycling bake curve as compared to that after cycling but without bake. Therefore, it is argued that the retention characteristics are fairly excellent.

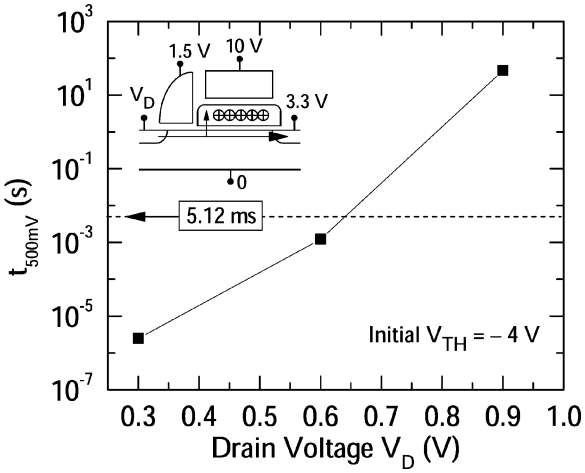

Fig. 13. Measured soft-write characteristic for the over-erased cells in the selected SG line during programming under worst-case conditions. $t_{500\ \mathrm{mV}}$ is the time to a V_{TH} shift of 500 mV. The disturb time for byte programming is 5.12 ms, which validates soft-write disturb inhibition at $V_D = 1$ V in this paper.

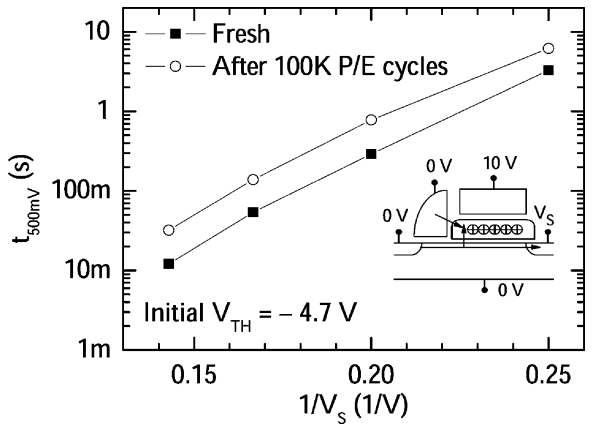

Fig. 14. Measured MB disturb characteristics as a function of V_S for cells before and after 10^5 P/E cycles. $t_{500\ \mathrm{mV}}$ is the time for a V_{TH} shift of 500 mV. The disturb time is 70 μs, which validates MB disturb immunity for the case of $V_S = 3.3$ V.

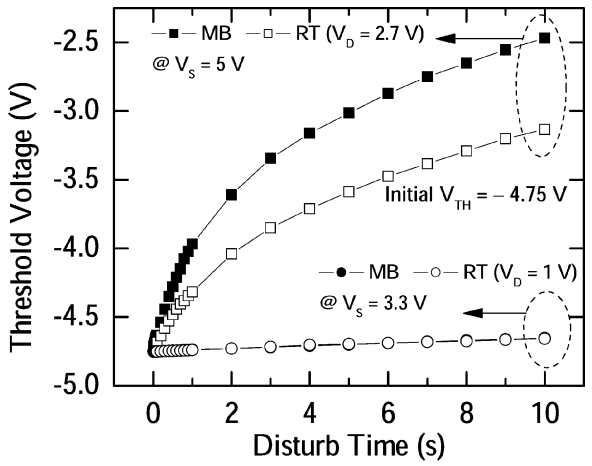

Fig. 15. Measured MB and RT disturb characteristics in terms of threshold voltage variation for $V_S = 3.3$ V and 5 V. Use of $V_S = 3.3$ V leads to MB and RT disturb immunity.

C. Disturb Effects

As mentioned in the Experimental section, the specific page arrangement can isolate disturbs from the other pages through CG and source switch; that is, an individual page is unlikely to be affected when other pages undergo P/E cycles. However, this is not the case at the cell level within single page, as explained below.

1) Program Inhibition: During programming, soft-write effect might occur when the B/L voltage fails to prevent the inhibited cells in selected W/L [e.g., C1 and C3 in Fig. 3(b)] from being programmed. Fig. 13 shows measured soft-write characteristic for over-erased cells under the worst-case conditions. The $t_{500\ \mathrm{mV}}$ in Fig. 13 is the time to a 500-mV V_{TH} shift. Considering byte programming with a programming time of 10 μs in a 4-Kb W/L, the worst case disturb time of 5.12 ms ($= 4096 \times 10\ \mu\mathrm{s}/8$) is drawn as labeled in Fig. 13. It can be seen that even for the cell erased to -4 V, a drain voltage of 1 V can effectively inhibit the undesired programming. This argument holds for the test samples in the work. At the mass production level, however, to compensate for possible variation in threshold voltage and channel length of SG devices, a larger V_D must be considered.

2) Program Disturbs: For deselected over-erased cells in the same B/L where there is a selected cell undergoing programming [e.g., C5 in Fig. 3(b)], the W/L voltage (V_{SG}) and B/L voltage (V_{D}) both are biased at 0 V and an insignificant channel current should be expected to prevent unwanted programming in these cells. However, due to select-gate channel length reduction or threshold voltage shift caused by process variations, the exponentially increased channel subthreshold current may erroneously induce electron injection from the split point onto floating gate (namely, the subthreshold source-side injection, S^3I [11]). Meanwhile, there may be another mechanism occurring, the reverse tunneling of electrons, from SG through interpoly oxide to FG. This is due to the over-erased state and the coupling of source voltage, making FG potential highly positive with respect to SG. On the other hand, one may consider the possible injection of electrons due to avalanche or impact generation at the split point in the presence of an inversion layer formed beneath floating gate. However, as will be explained later, we must rule out the possibility of avalanche or impact generation. Therefore, the so-called "mirror-bit (MB) disturb" as conducted at $V_D = 0$ V may involve subthreshold SSI (S^3I) and/or reverse tunneling (RT). Fig. 14 depicts the MB disturb data, in terms of the time to a 500-mV threshold voltage shift versus the reciprocal of V_S, measured from the over-erased cells before and after 10^5 P/E cycles. The physical origin of the data in this figure will be addressed in detail later. For a selected page under programming, each over-erased cells can only tolerate at most seven pulses of MB disturb from the selected cells in the same B/L. Thus, in the byte programming mode, the worst case MB disturb time is calculated to be 70 μs ($= 7 \times 10\ \mu$s), indicating the MB disturb immunity for the involved V_S of 3.3 or even 5 V as can be validated in Fig. 14.

It is interesting to further examine the case of $V_S = 3.3$ V. To suppress subthreshold SSI, an inhibition voltage $V_D = 1$ V is applied. The results are shown in Fig. 15 along with those from $V_{\mathrm{D}} = 0$ V for comparison. It can be seen that the threshold

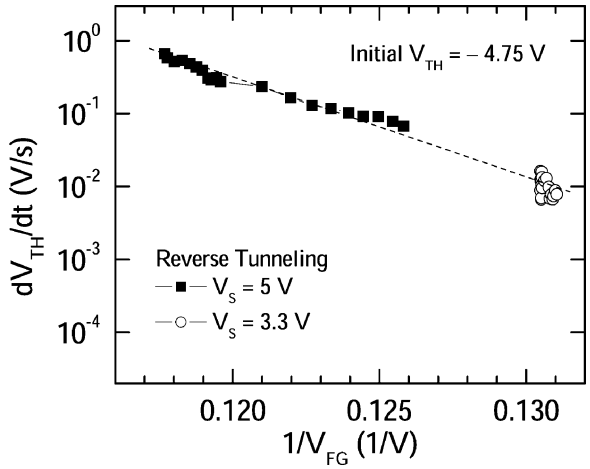

Fig. 16. Time derivative of threshold voltage versus reciprocal of floating gate voltage, transformed from RT data of $V_S = 5$ V and the data of $V_S = 3.3$ V in Fig. 15. A straight line fitting the RT data of $V_S = 5$ V is extended to enter into the $V_S = 3.3$ V regime.

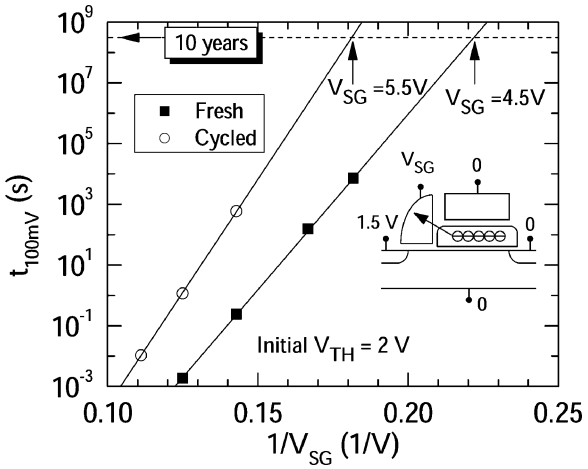

Fig. 17. Measured soft-erase characteristics under worst-case conditions during read-out operation before and after 10^5 P/E cycles. $t_{100\ \mathrm{mV}}$ is the time to a V_{TH} shift of 100 mV.

voltage slightly increases with increasing disturb time, regardless of $V_D = 0$ or 1 V. The physical origin is that the kinetic energy of hot electrons near the split point is not sufficiently large as compared with the SiO_2/Si barrier height of around 3.2 eV. Therefore, the reverse tunneling seems to prevail in this case. On the other hand, a higher inhibition voltage V_D of 2.7 V for effective suppression of subthreshold SSI is done for $V_S = 5$ V. The measurement results are again plotted in Fig. 15 along with those from $V_D = 0$ V. Obviously, both subthreshold SSI and reverse tunneling exist in the MB disturb measurement ($V_D = 0$ V) but with the former as the dominant factor. Further insights can be obtained by differentiating the threshold voltage of interest in Fig. 15 with respect to the disturb time [12]. The results are shown in Fig. 16 against floating gate voltage. Here the FG voltage is calculated by using a conversion formula [13]: $V_{\mathrm{FG}} = \alpha_G V_{\mathrm{CG}} + \alpha_{\mathrm{SG}} V_{\mathrm{SG}} + (\alpha_{\mathrm{SG}} + \alpha_S) V_S + Q_{\mathrm{FG}}/C_T$, where the α_i's, Q_{FG} and C_T are the coupling ratio of each terminal, the floating-gate charge, and the total capacitance, respectively. It is noteworthy that in the present cell structure, the importance of the fringing capacitive coupling from the source should not be absent, as detailed elsewhere [13]. In other words, in the presence of this fringing component, the source coupling in inversion conditions appears to be larger than CG coupling. It can be seen from Fig. 16 that the data of $V_S = 3.3$ V fall on or close to a straight line that fits the RT data of $V_S = 5$ V. This reasonably determines reverse tunneling dominating in the $V_S = 3.3$ case.

The aforementioned analysis dedicated to $V_S = 3.3$ and 5 V can be readily applied to elucidate measured data in Fig. 14 in a wide range of V_S. First of all, the values of V_S in Fig. 14 can essentially produce injected electrons with kinetic energy larger than SiO_2/Si barrier height of around 3.2 eV. It therefore is expected that an increase in V_S can produce a large (i.e., exponential) change in the amount of injected electrons via subthreshold SSI. At the same time, the floating gate voltage is increased via both inversion-layer coupling and fringing capacitive coupling, which in turn produces exponential increase in reverse tunneling. All of these two are consistent with experimental data in Fig. 14: a small increase in V_S gives rise to an exponential decrease in time to critical threshold voltage shift. On the other hand, we have found that for the V_S range involved in Fig. 14, the substrate current is limited by the background noise ($\sim 10^{-11}$ A), reasonably removing the possibility of avalanche or impact generation.

Note that to highlight the disturb mechanisms, both Figs. 14 and 15 use the same initial V_{TH} (in the deep over-erased state). Since the reverse tunneling takes place in the cells with inhibited V_D and zero V_{SG} in selected page [e.g., C4 and C6 in Fig. 3(b)], the worst case time is calculated to be 35.83 ms ($= 7 \times 4095 \times 10\ \mu\mathrm{s}/8$) for byte programming. Again returning back to Fig. 15, the cells experiencing disturbs are judged to have good tolerance.

3) Soft Erase: Soft erase may occur after repeated read-out operation on the programmed cell [6]. Such effect is significant in the case of electron loss through interpoly oxide [14]. Fig. 17 shows the worst case soft-erase characteristics measured on selected programmed cell during read-out operation, plotted before and after 10^5 P/E cycles. It can be seen that the projected V_{SG} of the cycled cell at the ten-year limit is higher than that of fresh cell. This indicates the presence of the trapped electrons in interpoly oxide after P/E cycles. A select-gate voltage of 4 V is therefore determined while meeting high read-out current and soft-erase immunity of more than ten years.

VI. Further Discussion

Since its appearance in 1980s, there have been a large number of articles dedicated to the split-gate Flash device using source-side hot-electron injection [1]–[3], [5], [6], [15]–[19]. Specifically, the latest sidewall split-gate NOR Flash as presented in this paper can be traced back to the triple-poly sidewall split-gate Flash memory in 1989 [2]. It is well recognized that due to extremely complicated triple-poly processes at that time, little information was heard afterwards about the mass production of the triple-poly sidewall split-gate Flash memory. Only with the significant progress in the manufacturing technology in the recent years can it become possible for realization of the triple-poly sidewall split-gate Flash memory in a CMOS compatible manufacturing line.

Indeed, our proposed cell structure exhibits apparent differences as compared with the existing source-side injection and/or split-gate structures. First, the rounded corner of the FG sidewall substantially improves the interpoly oxide reliability. Second, the novel device uses CG coupling rather than under-diffused source in other split-gate devices [15]. Therefore, the dimensional (width and length) down-scaling is promised due to suppression of punchthrough disturb [20]. Thirdly, due to the self-aligned fabrication of FG against CG and the subsequent self-aligned SG against CG, the number of the mask (and thus the process cost) can be greatly reduced even in the 0.15-μm technologies. On the contrary, the other split-gate devices were reported to add extra poly layers to a two-poly process [21], [22] for the purposes of self-alignment and poly-injector fabrication.

The proposed cell structure can find other potential applications in the NAND architecture. One of the examples is the 512-Mb sidewall split-gate NAND Flash memory in the 0.12-μm process [23]. In this NAND architecture, a fast and low-voltage (< 11 V) source-side hot-electron injection is employed for programming in size of pages because of high-injection efficiency of split-gate. Moreover, the same cell structure has again proved itself in the realization of the multilevel concept in the NAND architecture [24].

VII. Conclusion

A new sidewall split-gate 32-Mb Flash memory with the rounded FG sidewall corner has been fabricated in the NOR architecture. The endurance test has manifested the potential of the rounded FG sidewall. The program and read disturbs have been examined and the operating conditions for disturbs inhibition have readily been determined. We have also elaborated on the differences between the proposed cell structure and the existing ones, as well as on the NAND architecture application.

References

[1] A. T. Wu, T. Y. Chan, P. K. Ko, and C. Hu, "A novel high-speed, 5-volt programming EPROM structure," in *IEDM Tech. Dig.*, 1986, pp. 584–587.

[2] K. Naruke, S. Yamada, E. Obi, S. Taguchi, and M. Wada, "A new Flash erase EEPROM cell with a sidewall select-gate on its source side," in *IEDM Tech. Dig.*, 1989, pp. 603–606.

[3] C.-F. Chen, "Flash Memory Cell with Self-Aligned Gates and Fabrication Process," U.S. Patent 6 291 297 B1, Sep. 18, 2001.

[4] M. Branchetti, G. Campardo, S. Commodaro, S. Ghezzi, A. Ghilardelli, C. Golla, I. Martines, M. Maccarrone, R. Micheloni, M. Zammattio, and S. Zanardi *et al.*, "Memory architecture and related issues," in *Flash Memories*, P. Cappelletti *et al.*, Eds. Norwell, MA: Kluwer, 1999, pp. 264–266.

[5] J. Van Houdt, D. Wellekens, L. Faraone, L. Haspeslagh, and L. Deferm, "A 5 V-compatible Flash EEPROM cell with microsecond programming time for embedded memory applications," *IEEE Trans. Compon., Packag., Manufact. Technol. A*, vol. 17, pp. 380–389, Sep. 1994.

[6] D. Wellekens, J. Van Houdt, L. Faraone, G. Groeseneken, and H. E. Maes, "Write/erase degradation in source side injection Flash EEPROM's: Characterization techniques and wearout mechanisms," *IEEE Trans. Electron Dev.*, vol. 42, no. 11, pp. 1992–1998, Nov. 1995.

[7] C. Chen, Z.-Z. Liu, and T.-P. Ma, "Analysis of enhanced hot-carrier effects in scaled Flash memory devices," *IEEE Trans. Electron Devices*, vol. 45, no. 7, pp. 1524–1530, Jul. 1998.

[8] P. Cappelletti and A. Modelli *et al.*, "Flash memory reliability," in *Flash Memories*, P. Cappelletti *et al.*, Eds. Norwell, MA: Kluwer, 1999, pp. 417–418.

[9] W. H. Lee, D.-K. Lee, T.-M. Park, K.-S. Kim, K.-O. Ahn, and K.-D. Suh, "Data retention failure in NOR Flash memory cells," in *Proc. 2001 39th Int. Reliab. Phys. Symp.*, pp. 57–60.

[10] H.-T. Huang and M.-J. Chen, "A novel sphere-based statistical model for local oxide thinning induced gate oxide breakdown," in *Proc. 2000 Int. Conf. Solid State Devices and Mater.*, pp. 246–247.

[11] J. Van Houdt, L. Haspeslagh, L. Deferm, G. Groeseneken, and H. E. Maes, "Method of programming a Flash EEPROM memory cell array optimized for low power consumption," U.S. Patent 6 044 015, Mar. 28, 2000.

[12] J. Van Houdt, L. Haspeslagh, D. Wellekens, L. Deferm, G. Groeseneken, and H. E. Maes, "HIMOS—A high efficiency Flash $\mathrm{E^2PROM}$ cell for embedded memory applications," *IEEE Trans. Electron Devices*, vol. 40, no. 12, pp. 2255–2263, Dec. 1993.

[13] C. Y.-S. Cho and M.-J. Chen, "Improved subthreshold slope method for precise extraction of gate capacitive coupling coefficients in stacked gate and source-side injection flash memory cells," *Solid State Electron.*, vol. 48, pp. 1189–1195, 2004.

[14] G. Groeseneken and H. E. Maes, "A quantitative model for the conduction in oxides thermally grown from polycrystalline silicon," *IEEE Trans. Electron Devices*, vol. ED-33, no. 7, pp. 1028–1042, Jul. 1986.

[15] S. Kianian, A. Levi, D. Lee, and Y.-W. Hu, "A novel 3 volts-only, small sector erase, high density Flash $\mathrm{E^2PROM}$," in *Symp. VLSI Tech. Dig.*, 2000, pp. 71–72.

[16] J. V. Houdt, G. Groeseneken, and H. E. Maes, "An analytical model for the optimization of source-side injection flash EEPROM devices," *IEEE Trans. Electron Devices*, vol. 42, no. 10, pp. 1314–1320, Oct. 1995.

[17] Y. Ma, C. S. Pang, J. Pathak, S. C. Tsao, C. F. Chang, Y. Yamauchi, and M. Yoshimi, "A novel high density contactless flash memory array using split-gate source-side injection cell for 5 V-only application," in *Symp. VLSI Tech. Dig.*, vol. 5A, pp. 49–50.

[18] K.-C. Huang, Y.-K. Fang, D.-N. Yaung, C.-W. Chen, H.-C. Sung, D.-S. Kuo, C. S. Wang, and M.-S. Liang, "The impacts of control gate voltage on the cycling endurance of split gate flash memory," *IEEE Electron Device Lett.*, vol. 21, no. 4, pp. 359–361, Apr. 2000.

[19] H. Guan, D. Lee, and G. P. Li, "An analytical model for optimization of programming efficiency and uniformity of split gate source-side injection Superflash memory," *IEEE Trans. Electron Devices*, vol. 50, no. 5, pp. 809–815, May 2003.

[20] *IEEE standard definitions and characterization of floating gate semiconductor arrays*, IEEE Std 1005-1998, Feb. 9, 1999.

[21] G.-C. Chern, "Self aligned method of forming a semiconductor memory array of floating gate memory cells with floating gates having multiple sharp edges, and a memory array made thereby," U.S. Patent 6 750 090 B2, Jun. 15, 2004.

[22] G.-C. M. Chern and C.-S. Su, "Self aligned method of forming a semiconductor array of non-volatile memory cells," U.S. Patent 6 706 592 B2, Mar. 16, 2004.

[23] C.-Y. Hsu, C.-W. Hung, D. Sung, C.-S. Wu, S. C. Chen, H. H. Kuo, J. Y. Pan, C. L. Chen, I. C. Chuang, V. Huang, C. C. Hsue, D.-T. Fan, J.-C. Lu, C. Y.-S. Cho, K. Tseng, A. Hsu, B. Sheen, P. Tuntasood, and C.-F. Chen, "Split-gate NAND Flash memory at 120 nm technology node featuring fast programming and erase," in *Symp. VLSI Tech. Dig.*, 2004, pp. 78–79.

[24] H.-H. Kuo, J.-Y. Pan, C.-L. Chen, I.-C. Chuang, S.-C. Chen, C.-Y. Hsu, D. Sung, C.-W. Hung, V. Huang, C.-C. Cho, C.-C. Hsue, D.-T. Fan, J.-C. Lu, K. Tseng, J. Pabustan, S. Wang, A. Hsu, B. Sheen, P. Tuntasood, and C.-F. Chen, "Multi-level split gate NAND memory with fast program and erase in 120 nm," in *Proc. 2004 NVSMW*, pp. 37–39.

Caleb Yu-Sheng Cho (S'04-M'05) received the B.S. degree in electrical engineering from National Sun Yat-Sen University, Kaohsiung, Taiwan, R.O.C., the M.S. degree in electronics engineering from National Tsing Hua University, Hsinchu, Taiwan, R.O.C., and the Ph.D. degree in electronics engineering from National Chiao-Tung University, Hsinchu, in 1996, 1999, and 2004, respectively.

In 2004, he joined Actrans System, Inc., Hsinchu. His research interests are Flash memory device physics, coupling ratio extraction, novel Flash memory structure development, and Flash memory circuit design.

Ming-Jer Chen (S'78–M'79–SM'98) received the B.S. degree in electrical engineering (with highest honors) from National Cheng-Kung University, Tainan, Taiwan, R.O.C., and the Ph.D. degree in electronics engineering from National Chiao-Tung University (NCTU), Hsinchu, Taiwan, R.O.C., in 1977 and 1985, respectively.

In 1985, he joined the faculty of the Department of Electronics Engineering, NCTU, where he became Full Professor in 1993. From 1987 to 1992, he served as a Consultant for the Taiwan Semiconductor Manufacturing Company (TSMC), Hsinchu, where he led a team from NCTU and Electronics Research and Service Organization (ERSO/ITRI) to build up a series of process windows and design rules. He has been granted six U.S. patents and six Taiwanese patents covering the high-precision analog capacitors, 1-T SRAM cell, dynamic threshold MOS, and ESD protection. From 2000 to 2001, he was a Visiting Professor at the Department of Electrical Engineering and the Center for Integrated Systems, Stanford University, Stanford, CA. His research interests have long been focused on the technology reliability physics and currently on nanoscale electronics. He has graduated 12 Ph.D. students and more than 80 Master's students.

Dr. Chen is a member of Phi Tau Phi.

Chiou-Feng Chen received the B.S. degree in electrical engineering from National Chen-Kung University, Taiwan, R.O.C., and the M.S. and Ph.D. degrees in electronics engineering from National Chiao-Tung University, Hsinchu, Taiwan, R.O.C., in 1979, 1981, and 1986, respectively.

He has been working in the nonvolatile memory industry for 15 years. He is now with Actrans System, Inc., Santa Clara, CA.

Prateep Tuntasood received the B.Eng. degree in electrical engineering from Chulalongkorn University, Bangkok, Thailand, and the M.S. and Ph.D. degrees in electrical engineering from University of Cincinnati, Cincinnati, OH in 1976, 1979, and 1982, respectively.

He is currently a Senior Member of the Technical Team, Flash Memory Development Group, Actrans System, Inc., Santa Clara, CA. Before joining Actrans System, Inc., he has been working at several semiconductor companies, including NexFlash Technologies, Inc.

Der-Tsyr Fan received the B.S. degree in chemistry from National Tsing-Hua University, Hsinchu, Taiwan, R.O.C., and the Ph.D. degree in materials science and engineering from Lehigh University, Bethlehem, PA, in 1981 and 1991, respectively.

Since then, he has been working in the areas of compound semiconductor and VLSI processing. He joined Actrans System, Inc., Hsinchu, Taiwan, R.O.C., in 2000 focusing on Flash memory technology integration. He is currently Vice President of technology.

Tseng-Yi Liu received the B.S. degree in electrical engineering from Tamkang University, Taipei, Taiwan, R.O.C. in 1994, and the M.S. degree in electronics engineering from Yuan Ze University, Taoyuan, Taiwan, R.O.C. in 1994 and 1996, respectively.

From 1998, he worked for Macronix, Hsinchu, Taiwan, R.O.C. as a Flash Memory Design Engineer. He participated in the circuit design of stacked-gate NOR and NBit Flash memories as a Project Manager. He is currently the Design Manager and the Leader for split-gate Flash memory design in Actrans System, Inc., Hsinchu.