A Novel Erase Scheme to Suppress Overerasure in a Scaled 2-Bit Nitride Storage Flash Memory Cell

Chih-Chieh Yeh, Tahui Wang, *Senior Member, IEEE*, Wen-Jer Tsai, Tao-Cheng Lu, Yi-Ying Liao, Hung-Yueh Chen, Nian-Kai Zous, Wenchi Ting, Joseph Ku, and Chih-Yuan Lu, *Fellow, IEEE*

***Abstract*—The cause of over–erasure in a two-bit nitride storage Flash memory cell is investigated. Extra positive charges accumulated above the n^+ junction and channel-shortening enhanced drain-induced barrier lowering effect are found to be responsible for threshold voltage (V_t) lowering in an over-erased cell. A modified erase scheme is proposed to resolve this issue. By applying a source voltage during erase, the erase speed can be well controlled for cells with different channel lengths and a wide range of program-state V_t distribution, which will reduce overerasure significantly.**

***Index Terms*—Band-to-band hot hole, Flash memory cell, nitride trapping storage, overerasure.**

I. Introduction

NITRIDE-BASED localized trapping storage Flash memory cells [1]–[4] have been demonstrated to be a promising candidate for nanoscale Flash memories because of their simple process, smaller bit size, and absence of floating-gate coupling effect. In spite of its many advantages, reliability issues including overerasure, high-V_t state charge loss, and low-V_t state charge gain [4]–[7] are shown to be limiting factors for device scaling. Overerasure of conventional floating-gate memory cells has been studied for a long time. It comes from a fast-erased cell, whose erase speed is increased by intrinsic oxide trap/defect enhanced electron tunneling during erase operation [9]–[12]. Nitride storage memory is demonstrated to have good immunity to intrinsic oxide defects [1], [2]. However, overerasure is still observed [4], [7]. In this paper, the dominant mechanism of V_t lowering in an over-erased nitride Flash cell is proposed, which is completely different from the mechanism in a conventional floating-gate memory. Besides, a novel erase scheme is proposed to suppress the overerasure.

The samples used in this work are n-channel MOSFETs with an oxide–nitride–oxide (ONO) gate dielectric stack (Fig. 1). The thickness of each ONO layer is 9 (top oxide), 6 (nitride), and 6 nm (bottom oxide). The gate length ranges from 0.3 to 0.35 μm. The program is done by channel hot electron injection ($V_g = 9$ V, $V_d = 5$ V, and $V_s = 0$ V for the bit at the drain side) while erase is done by band-to-band hot hole injection ($V_g = -6$ V, $V_d = 5$ V, and $V_s = 0$ V for the bit at the drain side). The read operation is performed by a reverse read scheme to realize two-bit storage ($V_g = 3$ V, $V_d = 0$ V, and $V_s = 1.6$ V to read the bit at drain side). The cells are arranged in a virtual ground array [1]. To simplify the experiments, a 1-bit-per-cell operation is performed and the results are applicable to 2 bits per cell as well.

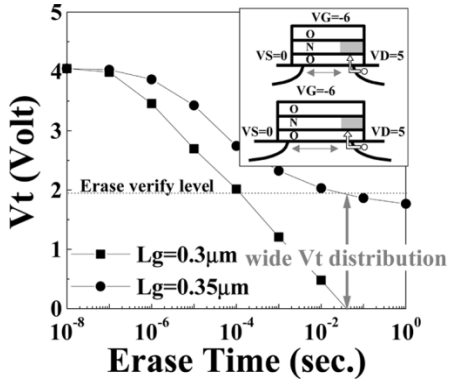

Fig. 1. Erase behavior of scaled nitride storage memory cells with gate lengths of 0.3 and 0.35 μm. The erase biases are $V_g = -6$ V, $V_d = 5$ V, and $V_s = 0$ V. The initial V_t of the 0.3- and 0.35-μm cell is 1.6 and 1.8 V, respectively. The threshold voltage (V_t) is defined as the applied gate voltage at which the drain current is 1 μA.

II. Erase Behavior and Charge Profiling in a Scaled Memory Cell

Fig. 1 shows the erase speed dependence on channel length. Shorter channel length ($L_g = 0.3$ μm) shows faster erase speed than the longer one ($L_g = 0.35$ μm) under the same erase voltages ($V_g = -6$ V, $V_d = 5$ V, $V_s = 0$ V). Accordingly, if process-induced channel length variation and sector operation architecture are taken into consideration, a large erase speed variation is expected, and a wide-erased V_t distribution results (see Fig. 1, L_g of 0.3 to 0.35 μm is used to emulate 10% process variation during mass production for a target channel length of 0.33 μm). Large array leakage current is often caused by the over-erased cell, which will cause read error and program failure in a memory array. The charge pumping technique is utilized to profile the charge distribution [8], [13] (see Fig. 2). A tail of charge pumping current ($I_{\rm cp}$) in the $V_{\rm gh}$ ranged from -1 to 1 V is observed in both long- (0.35 μm) and short- (0.3 μm) channel devices. Two devices show similar hole accumulation and distribution, which results in the same amount of channel shortening (the drawings in Fig. 1). Besides, the V_t measured at $V_d = 0.1$ V in both of the two devices is around 2 V, while the difference of the two device's V_t measured at $V_d = 1.6$ V is more than

Manuscript received May 20, 2004; revised June 21, 2004. The review of this letter was arranged by Editor A. Chatterjee.

C.-C. Yeh, T. Wang, and W.-J. Tsai are with the Macronix International Company Ltd., Hsinchu 300, Taiwan, R.O.C. They are also with Department of Electronics Engineering, National Chiao-Tung University, Hsinchu 300, Taiwan, R.O.C. (e-mail: ccyeh@mxic.com.tw).

T.-C. Lu, Y.-Y. Liao, H.-Y. Chen, N.-K. Zous, W. Ting, J. Ku, and C.-Y. Lu are with the Macronix International Company Ltd., Hsinchu 300, Taiwan, R.O.C.

Digital Object Identifier 10.1109/LED.2004.833596

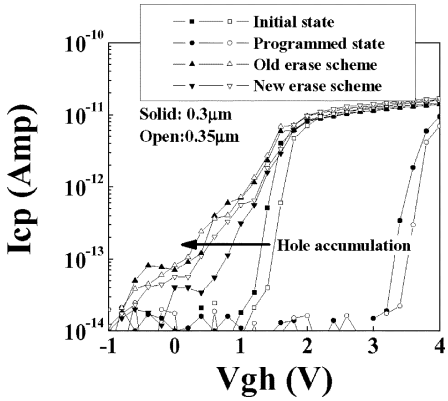

Fig. 2. Charge pumping characteristics of a nitride storage memory cells. The solid symbols denote the gate length of 0.3 μm, and the open symbols denote the gate length of 0.35 μm. The squares represent the initial behavior in a fresh cell. The circles represent the behavior in the program state. The up-triangles represent the behavior of the erased state with the old erase scheme ($V_g = -6$ V, $V_d = 5$ V, and $V_s = 0$ V) and an erase time of 10 ms, while the down-triangles represent the new erase scheme ($V_g = -6$, $V_d = 5$, and $V_s = 1$ V) and an erase time of 10 ms. Basically, the charge profile can be extracted from the experimental data from this figure with calculations [8], [13], although only raw data with qualitative analysis are provided here.

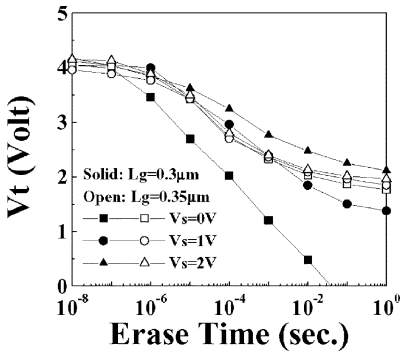

Fig. 3. Source bias and gate length effects on the erase speed. Three source biases of 0, 1, and 2 V and two gate lengths of 0.3 and 0.35 μm are used.

1 V. In long-channel devices, a little channel shortening will not affect the V_t much. However, channel shortening combined with the drain-induced barrier lowering (DIBL) effect in short-channel devices should dominate V_t reduction and overerasure.

III. Modified Erase Scheme to Suppress Overerasure

Fig. 3 shows the source bias (V_s) effect on erase speed of different gate lengths. The erase speed of a short-channel device shows a large dependence on V_s, while V_s has no effect on the erase speed of a long-channel device. In a short-channel device, V_s affects the electrical field at the drain side (hole injection side) significantly due to a shorter channel length. Higher V_s will reduce the lateral field at the drain side, thus reducing the hot hole injection efficiency. In contrast, in a long-channel device, V_s hardly affects the potential at the drain side and the erase speed has no dependence on V_s. Fig. 2 also shows the charge pumping current of an erased state by using the erase scheme with an applied V_s of 1 V. Less hole accumulation compared to the conventional scheme is observed in the short-channel device, while the longer one shows similar hole density at $V_s = 0$ V and $V_s = 1$ V. Accordingly, optimized V_s can reduce

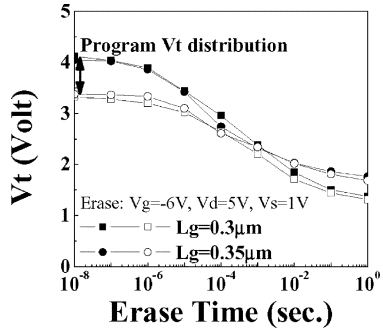

Fig. 4. Gate length and program-state V_t effects on the erase speed. Two programmed V_t's of 3.3 and 4 V and two gate lengths of 0.3 and 0.35 μm are used. The erase biases are $V_g = -6$ V, $V_d = 5$ V, and $V_s = 1$ V.

hole accumulation and overerasure in a short-channel device and, thus, alleviate the channel length deviation-induced erase speed variation. The impact of program-state V_t on erase speed is shown in Fig. 4. Two programmed V_t's of 3.3 and 4 V are used to emulate the high and low bound of a V_t distribution of 0.7 V in a memory array. As shown in the figure, both long- and short-channel devices with both high and low programmed V_t are erased and converged to about the same low V_t. The erase speed variation and overerasure induced by programmed V_t distribution and process variation can be suppressed by applying an optimized source bias (e.g., V_s is 1 V in our test sample).

According to these results, the modified erase scheme is recommended here for a scaled nitride storage memory. With an erase bias of negative V_g and positive V_d, small V_s is applied to control the erase speed for various channel lengths. An erase verify step is also needed to make sure if the cells have reached the desired level and to realize a tight erased V_t distribution.

IV. Conclusions

The DIBL, sector operation, and process variation will result in erase speed variation, which is the dominant cause of overerasure. A large erase speed dependence on source bias is observed in a shorter channel device, which can be utilized to modulate the erase speed. Despite a wide programmed V_t distribution and channel length variation, the final erased V_t is converged to the same voltage by applying a suitable erase source voltage.

References

[1] 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, pp. 543–545, Nov. 2000.

[2] M. K. Cho and D. M. Kim, "High performance SONOS memory cells free of drain turn-on and over-erase: Compatibility issue with current Flash technology," *IEEE Electron Device Lett.*, vol. 21, pp. 399–401, Aug. 2000.

[3] C. C. Yeh, W. J. Tsai, M. I. Liu, T. C. Lu, S. K. Cho, C. J. Lin, T. Wang, S. Pan, and C. Y. Lu, "PHINES: A novel low power program/erase, small pitch, 2-bit per cell Flash memory," in *IEDM Tech. Dig.*, 2002, pp. 37.4.1–37.4.4.

[4] T. Wang, W. J. Tsai, S. H. Gu, C. T. Chan, C. C. Yeh, N. K. Zous, T. C. Lu, S. Pan, and C. Y. Lu, "Reliability models of data retention and read-disturb in 2-bit nitride storage Flash memory cells (Invited)," in *IEDM Tech. Dig.*, 2003, pp. 7.4.1–7.4.4.

[5] 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. Non-Volatile Semiconductor Memory Workshop*, 2003, pp. 44–45.

[6] C. C. Yeh, W. J. Tsai, T. C. Lu, H. Y. Chen, H. C. Lai, N. K. Zous, G. D. You, S. K. Cho, C. C. Liu, F. S. Hsu, L. T. Huang, W. S. Chiang, C. J. Liu, C. F. Cheng, M. H. Chou, C. H. Chen, T. Wang, W. Ting, S. Pan, and C. Y. Lu, "Novel operation schemes to improve device reliability in a localized trapping storage SONOS-type Flash memory," in *IEDM Tech. Dig.*, 2003, pp. 7.5.1–7.5.4.

[7] W. J. Tsai, S. H. Gu, N. K. Zous, C. C. Yeh, C. C. Liu, C. H. Chen, T. Wang, S. Pan, and C. Y. Lu, "Cause of data retention loss in a nitride-based localized trapping storage Flash memory cell," in *Proc. IRPS*, 2001, pp. 34–38.

[8] C. Chen and T. P. Ma, "Direct lateral profiling of hot-carrier induced oxide charge and interface traps in thin MOSFETs," *IEEE Trans. Electron Devices*, vol. 45, p. 512, Feb. 1998.

[9] R. Shiner, J. Caywood, and B. Euzent, "Data retention in EPROMS," in *Proc. IRPS*, 1980, pp. 238–243.

[10] N. R. Mielke, "New EPROM data-loss mechanisms," in *Proc. IRPS*, 1983, pp. 106–113.

[11] G. Crisenza, G. Ghidini, S. Manzini, A. Modelli, and M. Tosi, "Charge loss in EPROM due to ion generation and transport in interlevel dieletric," in *IEDM Tech. Dig.*, 1990, pp. 107–110.

[12] F. Shone, H. Liou, C. Pan, B. Woo, and M. Hollwe, "A novel method to characterize and screen mobile ion contaminated nonvolatile memory products," in *Proc. VLSI Technology, Systems, and Applications*, 1991, pp. 224–226.

[13] M. Martirosian and T. P. Ma, "Lateral profiling of interface traps and oxide charge in MOSFET devices: Charge pumping versus DCIV," *IEEE Trans. Electron Devices*, vol. 48, pp. 2303–2309, Oct. 2001.