Performance and Reliability Evaluations of P-Channel Flash Memories with Different Programming Schemes

Steve S. Chung, S. N. Kuo, C. M. Yih, and T. S. Chao*

Department of Electronic Engineering, National Chiao Tung University, Taiwan, R.O.C.
** National Nano Device Lab., Hsinchu. Taiwan, R.O.C.*

Abstract

In this paper, a complete study of the cell reliability based on a unique oxide damage characterization for two different programming schemes of p-channel flash cell will be presented. These two programming schemes are Channel Hot Electron (CHE) injection or Band-to-Band (BTB) tunneling induced hot electron injection. Degradation of memory cells after P/E cycles due to the above oxide damages has been identified. It was found that both N_{it} and Q_{OX} will dominate the device degradation during programming. Although p-flash cell has high speed performance by comparing with n-flash cell, extra efforts are needed for designing reliable p-channel flash cell by appropriate drain engineering or related device optimization.

1. Introduction

Owing to its low voltage, low power and high speed programming features, p-channel flash memories [1-2] have been evolved as a promising cell for real applications in the future. In a certain design of p-channel flash cells, programming of the cell can be achieved either by channel-hot-hole impact ionization induced channel-hot-electron injection (CHE) or by band-to-band (BTB) tunneling induced hot-electron injection at the drain side. Erasing of the cell can be accomplished by electron channel Fowler-Nordheim (F-N) ejection from the floating gate. Here, both programming schemes, CHE and BTB, will generate the so-called oxide damages, which include the interface state N_{it} and the oxide trap charge Q_{ox}. These damages will give rise to serious reliability problems such as programming time delay, operation window closure, and data retention etc. for p-channel flash memories. In the past, the correlation of these program induced damages with device performance and reliability has been reported [3-4] and completely studied [5] for n-channel flash memory cells. However, thus far, it has never been reported for p-channel flash memories with different program schemes and so it is still not clear which programming scheme has better performance and reliability.

In this paper, for the first time, a comparative study of p-channel flash memory performance and reliability will be presented in details for two different programming schemes. The understanding of programming characteristics degradation mechanism for these two different programming methods will be made based a unique oxide damage characterization technique that we developed in [5]. Moreover, the comparison of reliability between n- and p-channel flash memory cells after long term P/E (program/erase) cycles will also be discussed.

2. Measurement and Characterization

Conventional double-polysilicon stacked-gate p-channel flash cells were used in this study. The tunnel oxide thickness is 70Å. The effective interpoly dielectric thickness with ONO structure is about 200Å. The measured gate coupling ratio of flash cell is about 0.6. In addition, the p-MOSFET dummy cell for oxide damage characterization has same tunnel oxide thickness, drawn channel length 0.55 μm, and width 100 μm. Performance and reliability of two different program schemes shown in Fig. 1 are evaluated. The hot carrier stress conditions were performed for the p-MOSFET dummy cells, i.e., the channel-hot-electron (CHE) stress at the maximum gate current ($I_{G,max}$) bias and the BTB stress at the bias with maximum injection efficiency.

The gate currents and injection efficiencies (I_G/I_{SUB} for CHE and I_G/I_D for BTB) of two programming schemes are shown in Figs. 2 and 3, for comparison. For CHE programming, the maximum gate current and the maximum injection efficiency are obtained at a gate voltage around the device threshold voltage. On the other hand, for BTB scheme, the gate current increases with increasing V_{GD} and the maximum injection efficiency is located at V_D= -6V and V_G= 4V. Here, we see that BTB scheme shows one order higher gate current and injection efficiency as compared to that of CHE. In Fig. 4, comparison of the transient characteristics of p-channel flash cells for CHE and BTB schemes is shown. It reveals that BTB scheme exihibits higher programming speed at the same drain bias than that of CHE. Fig. 5 gives a comparison of the measured gate current degradation with stress time for CHE and BTB schemes. It was found that BTB scheme degrades faster than the CHE one. This is consistent with the transient characteristics shown in Fig. 4. In short, BTB scheme has the advantage of high speed performance but with the drawback of a poorer reliability since BTB will generate larger oxide damages described as follows.

3. Results and Discussions

A. Characterization of Oxide Damages N_{it} and Q_{ox}

The oxide damage characterization method that we developed in [5] as also illustrated in Fig. 6 is used. The *gated diode currents* in combination with a special *detrapping tech-*

11.7.1

0-7803-4100-7/97/$10.00 © 1997 IEEE

nique is used for the determination of oxide damage lateral distributions. Once we have the oxide damages, memory cell performance and the cell degradation mechanisms can be fully explored. It is known that when the gated-diode measurement technique was first developed, it can only be used to characterize N_{it}. However, the localized damages of N_{it} and Q_{ox} exist simultaneously for both stressed conditions at BTB and CHE stress biases. Also the measured I_{GD} will be affected significantly by the created oxide charge as illustrated in Fig. 6(b). Therefore, we use the measured I_{GD} before and after the stress and develop a method to separate Q_{ox} from N_{it} as the following steps: (1) Measure the gated-diode (I_{GD}) current for a fresh p-MOSFET dummy cell (curve (1)). (2) After hot-electron stress, we have I_{GD} (curve (2)), which includes both N_{it} and Q_{ox}. (3) Use a neutralization step (or detrapping step) (@V_D= 2.5V, V_G= -3V for 50 sec in this study for CHE stress) to eliminate the effects of hot-electron injection induced Q_{ox}, (curve (3)). The difference between curves (1) and (3) gives N_{it}, while the difference between curves (2) and (3) gives Q_{ox}. Not only the experimental procedure as above, but also simulation are required to accurately determine the spatial distributions of N_{it} and Q_{ox}. Profiling of N_{it} and Q_{ox} can be obtained from equations listed in Table 1. In Fig. 6(a) for gated-diode current measurement configuration, a small drain bias is applied to forward bias the drain-substrate junction. It makes sure that the measured I_{GD} contributes to the electron and hole recombination process. The parameters Φ_e and Φ_h represent the quasi-Fermi levels for electron and hole respectively which coincide with the intrinsic level E_i. Δx between the locations of Φ_e and Φ_h is the overlap region where electron-hole recombination occurs and can be obtained from simulation. Fig. 7. gives the measured device gated-diode and GIDL currents for fresh, after CHE stress, and after the detrapping. Here, it should be noted that in the measurement as above, GIDL currents are used as a monitor of the detrapping procedure to make sure that Q_{ox} is eliminated in step (3) such that the detrapping step will not generate additional oxide damages.

For the cell after long term CHE programming in Fig. 1(a), the extracted lateral distributions of both $N_{it}(x)$ and $Q_{ox}(x)$ are illustrated in Fig. 8 (stressed @V_{GS}=1.1V and V_{DS}=-6V for 3000s). It was found that both $N_{it}(x)$ and $Q_{ox}(x)$ are localized inside the gate-drain overlap region. By taking the extracted oxide damages into account, the simulated programming characteristics of p-channel flash cells with and without oxide damages can be obtained as given in Fig. 9. Retardation of programming time is observed at a constant threshold voltage level. Moreover, the individual contribution of generated N_{it} and Q_{ox} to the programming characteristics is also shown in Fig. 9. With the same programming pulse, we see that the stress generated N_{it} will dominate the program window closure after long term CHE programming. For the cell after long term BTB programming, the extracted distributions of $N_{it}(x)$ and $Q_{ox}(x)$ are shown in Fig. 10 (stress @V_G= 4V and V_D=-6V for 3000s). From Fig. 10, we see that the oxide damages are much larger than those generated using CHE scheme shown in Fig. 8. It further shows in Fig. 11 that the induced oxide trap charges and interface states significantly and seriously affect the programming speed. The contribution of N_{it} to programmed threshold voltage variation is larger than that of Q_{ox} although the generated Q_{ox} is greater than N_{it}. However, the BTB scheme has one advantage of larger operation window when comparing Figs. 9 and 11.

B. Comparison of the Cell Reliabilities for CHE and BTB Programming Schemes

To compare the performance of a p-channel flash cell with different programming schemes, several results are given as follows. Although BTB scheme has poor gate current degradation characteristics, optimization of this degradation can be achieved by varying the BTB biases as shown in Fig. 12 with fixed V_{GD}=10V. In Fig. 12, we see that the smaller the drain bias, the lower the device gate current degradation. This implies that better reliability can be achieved by lowering the BTB drain bias. Further results given in Fig. 13 manifests that device with lower drain bias for BTB has better reliability. In other words, the programming speed depends largely either on interface states or on oxide trap charges. Therefore, a trade-off between programming speed and reliability for p-channel flash cells with BTB program should be considered. Furthermore, the cell programming time as a function of P/E cycles are compared in Fig. 14 among n- and p-channel flash cells with different programming schemes. It shows that BTB has the fastest programming speed before cycling but has the worst time delay as a result of the largest oxide damages, while the delay is moderate for a CHE scheme in p-channel flash cell. Also, one interesting result is that n-channel flash cell is much more reliable in this comparison. Initially, p-channel cell operates with faster speed but its speed will slow down due to the oxide damages after long term P/E cycles. Therefore, we conclude that the oxide damages, in particular N_{it}, which are the dominant degradation factor, should be suppressed either by drain-engineering or device optimization. A summary for the comparison of the p-flash cell performance is listed in Table 2.

4. Conclusions

In summary, several new issues have been addressed in this study: (1) Degradation of p-channel flash memory cells after P/E cycles due to oxide damages has been demonstrated. (2) It was found that both N_{it} and Q_{ox} will dominate the device degradation during programming. (3) By comparing with n-channel flash cell, extra efforts are needed for improving p-channel flash cell design by drain-engineering or related device optimization, and (4) We provide a methodology for evaluating cell reliability of either p-channel or n-channel flash memory. This provides us useful information for designing high performance and reliable flash memories.

Acknowledgments The authors would like to appreciate the financial support from the National Science Council, Taiwan, R.O.C. under grant *NSC85-2215-E009-053*.

References

[1] C. C.-H. Hsu et al., ***Ext. Abs. SSDM***, Tsukuba, p. 140, 1992.
[2] T. Ohnakado et al., in ***IEDM Tech. Dig.***, p. 279, 1995.
[3] Jack Z. Peng et al., ***Proceedings of IRPS***, p. 154, 1994.
[4] S. Yamada et al., in ***IEDM Tech. Digest***, p. 23, 1993.
[5] S. S. Chung et al., in ***Tech. Dig. Symp. VLSI Technology***, June 6-12, Kyoto, Japan, p. 111, 1997.

$$\Delta I_{GD}(V_G) = \frac{1}{2}\, qW\Delta S(x)\Delta x(V_G) n_i exp\left(\frac{q|V_D|}{2kT}\right) \quad (1)$$

$$\Delta I_{GD}(V_G) = \frac{1}{2}\, qW\Delta S(x)\Delta x\left(V_G + \frac{qQ_{ox}(x)}{C_{ox}}\right) n_i exp\left(\frac{q|V_D|}{2kT}\right) \quad (2)$$

$$\Delta S(x) = \sigma v_{th} \Delta N_{it}(x) \quad (3)$$

Table 1 Gated-diode current model for profiling N_{it} and Q_{ox}.

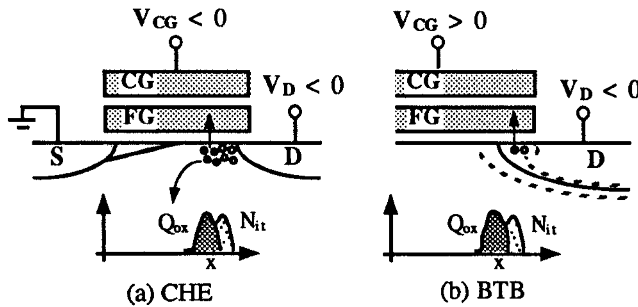

Fig. 1 The schematic illustration of two different programming schemes of a p-channel flash cell. Channel hot Electron (CHE) programming and Band-to-band (BTB) tunneling induced hot electron programming.

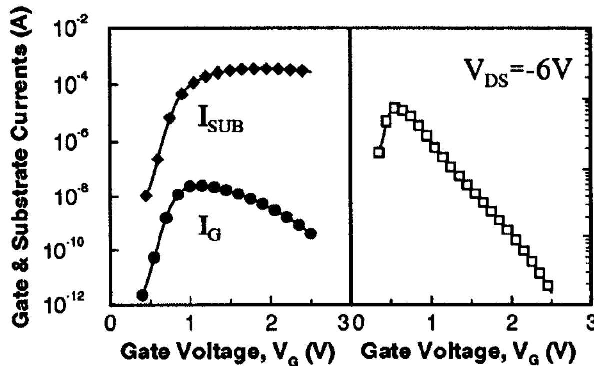

Fig. 2 The gate and substrate currents at V_{DS}= -6V and the calculated injection efficiency for CHE programming.

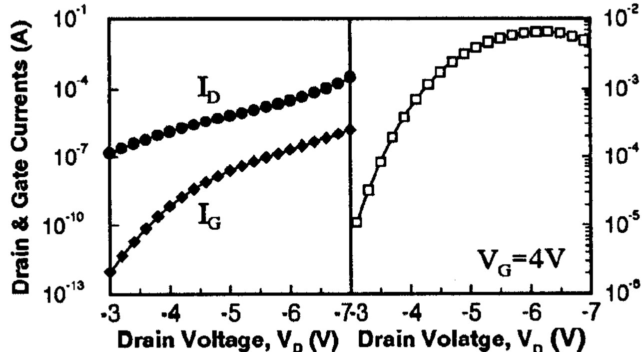

Fig. 3 The gate and drain currents at a fixed gate bias and the calculated injection efficiency for BTB programming.

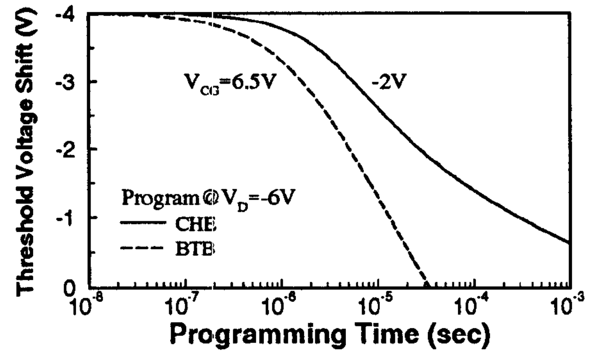

Fig. 4 Comparison of the transient characteristics for CHE programming and BTB programming.

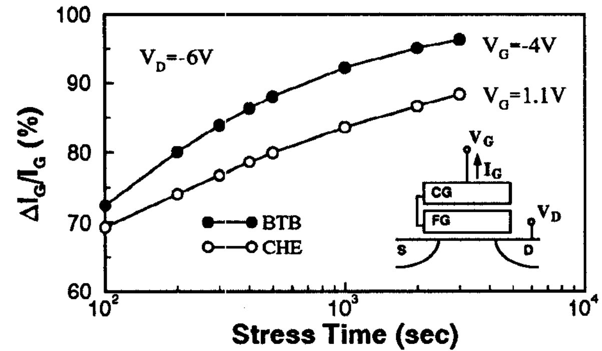

Fig. 5 The measured gate current degradation with stress time for CHE programming and BTB programming.

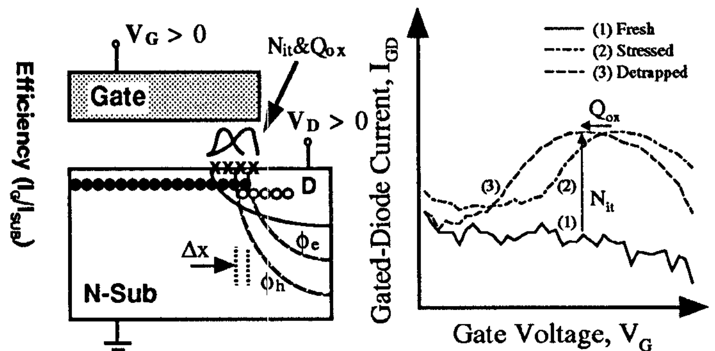

Fig. 6 (a) Schematic diagram for gated-diode measurement.
(b) The currents used for the calculation of N_{it} and Q_{ox}.

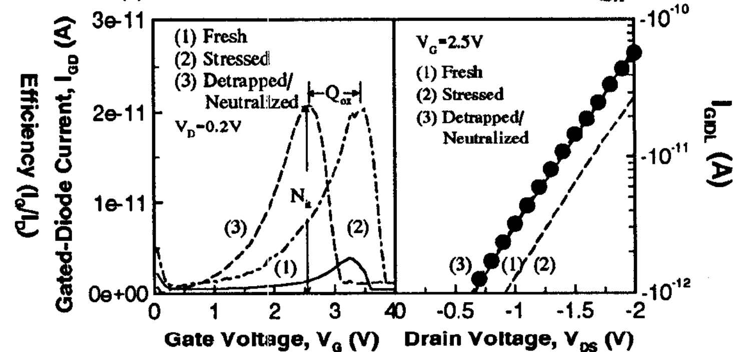

Fig. 7 The measured device gated-diode (left) and GIDL currents (right) for fresh, after $I_{G,max}$ stress (CHE, Hot-Electron stress) and after the detrapping.

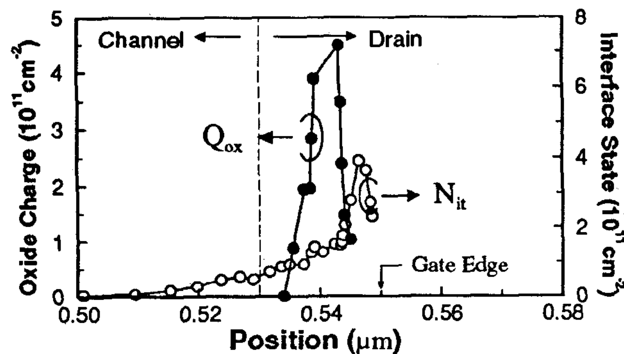

Fig. 8 Extracted lateral distributions of N_t and Q_x for devices with CHE stress at $I_{G,max}$ for 3000sec.

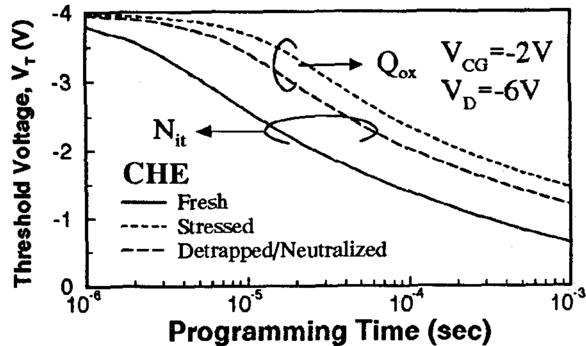

Fig. 9 The simulated writing characteristics of a flash cell during the fresh, after the stress and after detrapping steps, in which N_{it} dominates the device degradation.

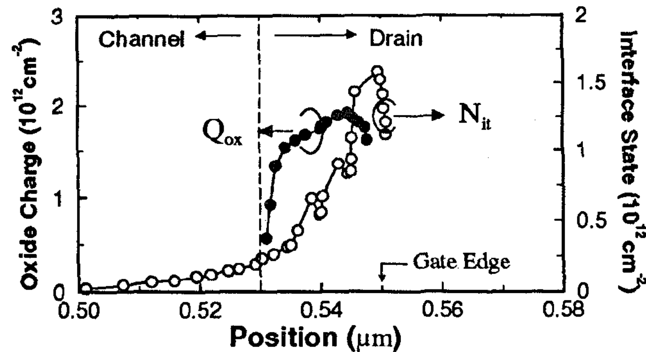

Fig. 10 Extracted lateral distributions of N_{it} and Q_{ox} for devices with BTB stress at the maximum injection bias for 3000 sec. in Fig. 3.

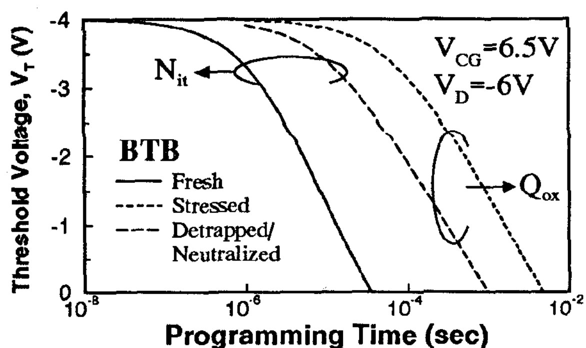

Fig. 11 The simulated writing characteristics of a flash cell during the fresh, after the stress and after detrapping steps, in which N_{it} dominates the device degradation.

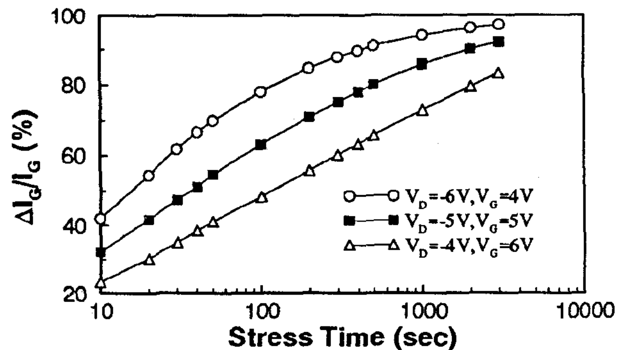

Fig. 12 Comparison of gate current degradations for BTB stress with different drain-gate bias programming.

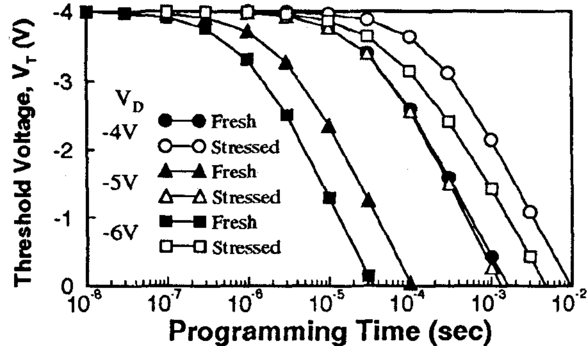

Fig. 13 Comparison of the transient characteristics for the various BTB stress conditions given in Fig. 12.

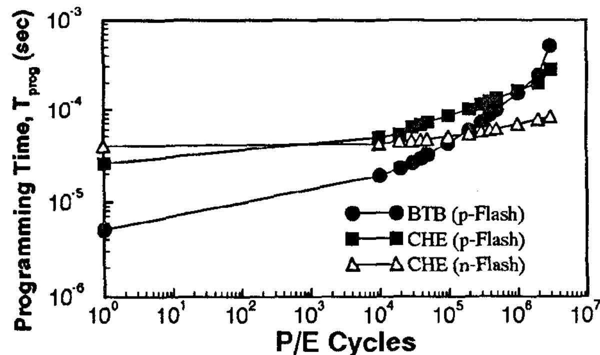

Fig. 14 Programming speed after P/E cycles for p-channel cell with BTB and CHE schemes, and for n-channel cell with CHE schemes.

Characteristics \ Programming Scheme	*BTB*	*CHE*
Injection Efficiency	Large	Small
Operation Window	Large	Small
$\Delta I_G/I_G$	Serious	Moderate
Programming Speed	Fast	Slow
Speed Retardation	Serious	Moderate

Table 2 Comparison of the performance for p-channel flash cells with different programming schemes.