Extended Abstracts of the 2016 International Conference on Solid State Devices and Materials, Tsukuba, 2016, pp107-108

B-6-01

47% Data-Retention Error Reduction of TLC NAND Flash Memory by Introducing Stress Relaxation Period with Round-Robin Wear-leveling

Yoshiaki Deguchi, Atsuro Kobayashi and Ken Takeuchi
Chuo University, 1-13-27 Kasuga, Bunkyo-ku, Tokyo, 112-8551 Japan,
Phone: +81-3-3817-7374, E-mail: deguchi@takeuchi-lab.org

Abstract

This study analyzes the influence of the interval of time between write/erase endurance stress and programming the final data for the data-retention evaluation in 1Xnm TLC NAND flash memories. During the interval of time after the write/erase endurance stresses, electrons are de-trapped from the tunnel oxide and eventually decreases the data-retention errors. By intentionally introducing the stress relaxation period, e.g. 3 hours, with round-robin wear-leveling, the data-retention error can be decreased by 47%.

1. Introduction

NAND flash memory is one of the most common non-volatile memories for widely various purposes because of the fast speed and the high density. However, in triple-level cell (TLC) NAND flash, the narrower threshold voltage (V_{TH}) margin increases data-retention error as shown in Fig. 1(a) [1]. To reduce data-retention errors, dynamic characteristics of V_{TH} should be evaluated to develop advanced error-correcting codes (ECCs) [2-3].

As the primary failure of NAND flash, data-retention error is caused by the stress-induced leakage current from the floating gate or the charge de-trap from the tunnel oxide [4]. The data-retention errors increase as the write/erase (W/E) cycles and as the data-retention time. This work presents for the first time that the data-retention error strongly depends on the interval of time between W/E endurance stress and programming data ($t_{S\text{-}P}$) as shown in Fig. 1(b). Then, this paper proposes to introduce the intentional relaxation period by th round-robin wear-leveling, shown in Fig. 1(c), which decreases data-retention errors by introducing the long $t_{S\text{-}P}$ over ten minutes. The wear-leveling is implemented in the NAND controller in SSDs and also work as averaging the W/E cycles of each NAND flash block [5]

2. Measurement Results and Discussions

The measurement method is described in Fig. 2. W/E endurance stress is applied to one block of 1Xnm TLC NAND flash without a break. Next, $t_{S\text{-}P}$ is the wait time or relaxation period before programming the final data for the data-retention evaluation. Then, the data-retention errors are evaluated as a function of the retention time, $t_{D.R.}$. During the relaxation period, $t_{S\text{-}P}$, trapped electrons in the tunnel oxide are gradually de-trapped and decrease the data-retention errors, as will be discussed below.

Fig. 3 shows the measured bit error rate (BER) of program-disturb errors as a function of $t_{S\text{-}P}$. The program-disturb errors do not depend on $t_{S\text{-}P}$, probably because Incremental Step Pulse Programming (ISPP) [6] precisely control the V_{TH}, irrespective of the amount of trapped electrons in the tunnel oxide or of the initial V_{TH} before programming the data. On the other hand, this paper reports for the first time that data-retention errors strongly depend on $t_{S\text{-}P}$ as shown in Fig. 4. When $t_{S\text{-}P}$ is 12 hours, BER decreases by 61%. If $t_{S\text{-}P}$ is short, the slope of BER increase is steep even at large $t_{D.R.}$. Fig. 5 shows the measured BER as a function of $t_{S\text{-}P} + t_{D.R.}$. Figs. 4-5 show that the data-retention errors are determined mainly by the relaxation period, $t_{S\text{-}P}$. Slopes of the measured BER increase are shown in Fig. 6. At the longer $t_{S\text{-}P}$, the relaxation effects caused maybe by the de-trapping decreases the data-retention errors. Slope of BER increase is decreased by 53% and 72% when $t_{S\text{-}P}$ is 3 hours and 24 hours, respectively.

Fig. 7 describes the measured V_{TH} distributions as a function of $t_{S\text{-}P}$ to evaluate data-retention errors. In Fig. 7(a), measured V_{TH} shift with (a) $t_{S\text{-}P}$ = 0 hour is much larger than the V_{TH} shift with (b) $t_{S\text{-}P}$ = 24 hours. On the other hand, if the different $t_{D.R.}$ is compared, the measured V_{TH} shift is as large as 0.13V from $t_{D.R.}$= 0day to 1day. However, the V_{TH} shift from $t_{D.R.}$= 1day to 2days is small. These results show the measured V_{TH} shift strongly depends on the stress relaxation time, $t_{S\text{-}P}$ as well as the data-retention time, $t_{D.R.}$. The average V_{TH} shifts during a day are shown in Fig. 8. Fig. 8(a) shows the average V_{TH} shift when $t_{D.R.}$ changes from zero to one day. In this figure, relaxation effects by introducing the relaxation period, $t_{S\text{-}P}$, is larger than Fig. 8(b) whose result is evaluated when $t_{D.R.}$ changes from one to two days. These results show that the measured V_{TH} shift are small in the short $t_{S\text{-}P}$ even at large $t_{D.R.}$. Fig. 9 shows the tail bit characteristics of the highest V_{TH} state, "G", when $t_{D.R.}$ is one day. When $t_{S\text{-}P}$ is long enough such as 24 hours, the tail bits decrease by 84% compared when $t_{S\text{-}P}$ is zero hour. If number of tail bits which is near the read reference voltage (V_{Ref}) is large, it will increase the BER because these bits are most likely to fail during the data-retention.

Fig. 10 shows the measured slope of BER increase by intentionally introducing the relaxation period by the round-robin wear-leveling. In the wear-leveling, data are evenly programmed to all blocks in the round-robin order so that the number of W/E cycles of each block becomes the same. That is, the programming are performed from Block1, Block2…Block 1024. As the number of blocks becomes larger, that is the memory capacity is larger, the same block is programmed in the longer interval of time. That is, the relaxation period is automatically inserted. In Fig. 10, slope of BER increase and $t_{S\text{-}P}$ are evaluated assuming that the number of pages is 384 which is 128 word lines and the access latency is shown in [7]. It is assumed that the W/E operation of NAND flash are continuously operated. This is the pessimistic and worst case scenario because in the real applications, the read or stand-by operation can be inserted between each W/E operation. As shown in Fig. 10, the wear-leveling improves the worst-case of BER increase by 54%, which corresponds to decreasing BER by 47% when $t_{D.R.}$ is 10 days.

3. Conclusions

In this work, the relaxation effects of the program stress on the data-retention errors are evaluated. By intentionally introducing the relaxation time between W/E cycles with the round-robin wear-leveling, the data-retention errors decrease by 47%.

References

[1] S. Tanakamaru et al., *IRPS*, pp. 3B. 3.1-3B. 3.6, 2013.
[2] Y. Cai et al., *HPCA*, pp. 551-563, 2015.
[3] S. Tanakamaru et al., *ASP-DAC*, pp. 83-84, 2013.
[4] K. Lee et al., *IEEE Trans. Electron Devices*, pp. 659-667, 2016.
[5] K. Takeuchi, *ISSCC*, Tutorial T-7, 2008.
[6] K-D. Suh, et al., *Dig. Tech. Papers, ISSCC*, pp.128-129, 1995.
[7] D. Sharma, presented at Flash Memory Summit, 2014.

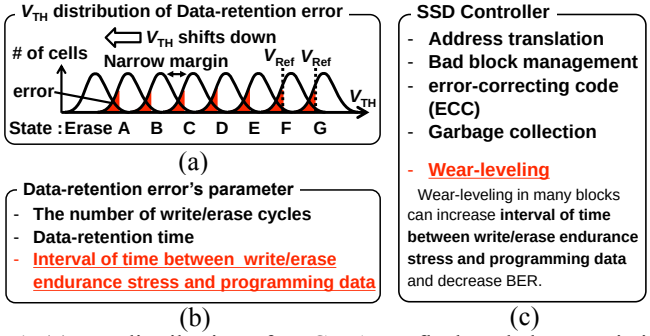

Fig. 1. (a) V_{TH} distribution of TLC NAND flash and characteristics of data-retention error. (b) New observation of this work. Data-retention error depends on the interval of time between write/erase (W/E) endurance stress and programming data ($t_{S\text{-}P}$) and can be decreased by (c) wear-leveling with round-robin order.

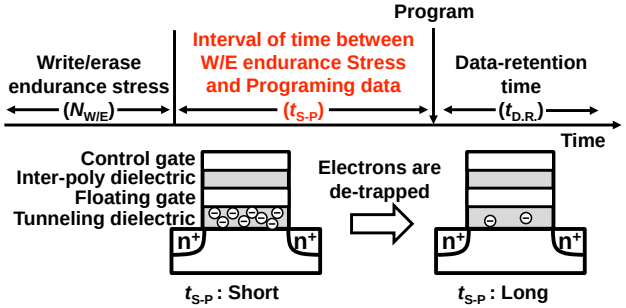

Fig. 2. Measurement flow of this work. When $t_{S\text{-}P}$ is short, many electrons stay trapped in tunneling dielectric. Trapped electrons are de-trapped at longer $t_{S\text{-}P}$. After programming data, data is read to evaluate the data-retention errors.

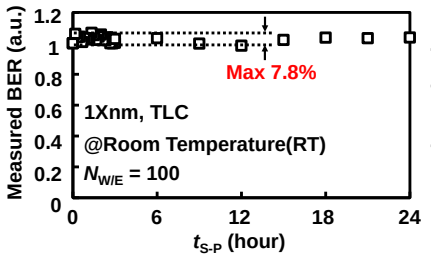

Fig. 3. Measured BER right after programming data vs. $t_{S\text{-}P}$. Even if $t_{S\text{-}P}$ changes, program-disturb error changes only by 7.8%.

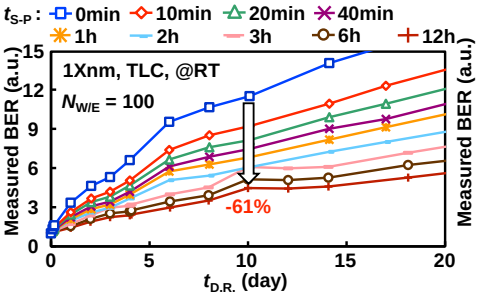

Fig. 4. New observation of the relaxation effect of data-retention errors. If $t_{S\text{-}P}$ is 12 hours, BER decrease by 61% compared with results whose $t_{S\text{-}P}$ is 0 hour.

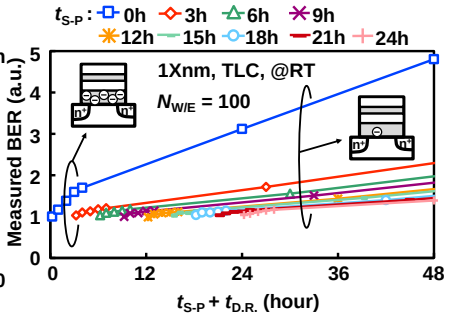

Fig. 5. Measured data-retention error vs. $t_{S\text{-}P}$ + $t_{D.R.}$. Slope of BER increase remains the same even if $t_{D.R.}$ is long.

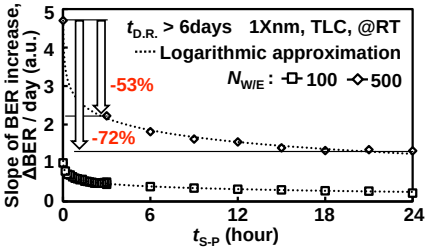

Fig. 6. Slope of BER increase per day vs. $t_{S\text{-}P}$. Logarithmic approximation fits well with the experimental results.

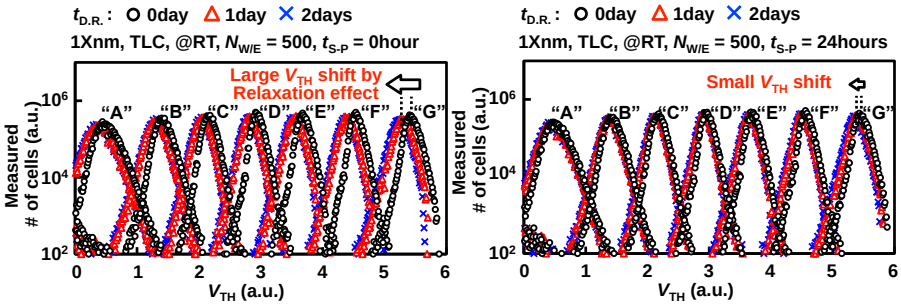

Fig. 7. Comparison of measured V_{TH} distributions in data-retention, (a) measurements when $t_{S\text{-}P}$ is zero hour and (b) measurements when $t_{S\text{-}P}$ is 24 hours. V_{TH} decreases significantly in case of (a), compared with (b). The longer relaxation time, $t_{S\text{-}P}$, suppresses the V_{TH} shift and decreases the data-retention errors.

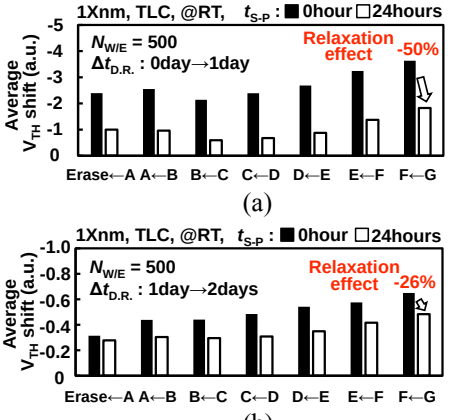

Fig. 8. Comparison of average V_{TH} shifts (a) when $t_{D.R.}$ is from zero day to one day, (b) when $t_{D.R.}$ is from one day to two days.

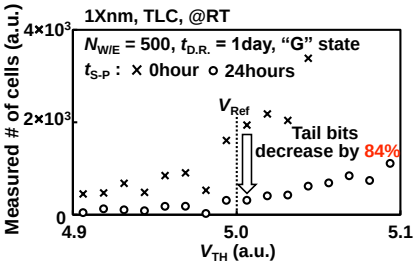

Fig. 9. Detailed tail bit of "G" state when $t_{D.R.}$ is one day. If $t_{S\text{-}P}$ is long, the number of cells near the read reference voltage which most likely becomes error decreases by 84%.

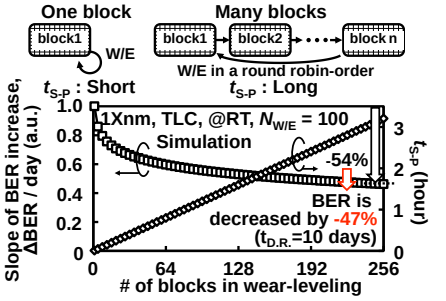

Fig. 10. Calculated slope of BER increase vs. number of blocks in the round-robin wear-leveling. By intentionally introducing the relaxation period by the wear-leveling, the data-retention error decreases by 47%.