Monitoring Flash EEPROM Reliability by Equivalent Cell Analysis

Daniele Ielmini
DEI-Politecnico di Milano
piazza L. da Vinci 32
20133 Milano - Italy
Tel. +39 02 2399 6120
ielmini@elet.polimi.it

Alessandro S. Spinelli
Dip. SSCCFFMM-Università
degli Studi dell'Insubria e INFM
via Valleggio 11 - Como - Italy
Tel. +39 031 238 6223
spinelli@uninsubria.it

Andrea L. Lacaita,
Massimiliano Gubello
DEI-Politecnico di Milano
piazza L. da Vinci 32
20133 Milano - Italy
Tel. +39 02 2399 6117
lacaita@elet.polimi.it

Michiel J. van Duuren
Philips Research Leuven
Kapeldreef 75
3001 Leuven - Belgium
Tel. +32 16 28 8539
michiel.van.duuren@philips.com

Abstract

The EC analysis allows to extract the distribution of stress-induced leakage current (SILC) in the memory array, without need for an extensive measurement of gate currents in the cells. In this paper, by comparing SILC distributions obtained by the EC concept and the real distributions, we demonstrate the accuracy of the new technique for assessing the reliability of non-volatile memories. The analysis for a Flash array with ultrathin oxide is finally carried out, showing how the EC method, in combination with a model accounting for the thickness dependence of SILC, provides a straightforward tool for evaluating the oxide degradation after cycling.

1. Introduction

The stress-induced leakage current (SILC) is a major issue affecting the data retention of floating-gate nonvolatile memories (NVMs). The increase of SILC for decreasing thickness of the tunnel oxide leads to enhanced tail effects during readout of cells, thus limiting the device scaling [1]. For this reason, the SILC has been investigated by several authors, usually by means of gate-current measurements on large area MOS capacitors. On the other hand, relatively few efforts have been devoted to the direct analysis of excess leakage currents in NVMs.

The standard technique for measuring the oxide leakage currents in small area NVM is the *gate stress*, namely a room-temperature retention experiment, accelerated by gate biasing [2]. From the measured time evolution of threshold voltage V_T, the gate current-voltage ($I-V$) characteristics of the cell can be evaluated [2, 3]. Based on the gate-stress technique, a new method for extracting the statistical distribution of SILC in NVM arrays was developed [4], which allows for a non-ambiguous characterization of memory reliability. Within this framework, the SILC distribution is extracted using the equivalent cell (EC) concept, which provides a fast scheme for reliability estimation of arrays in the Mbit range. The EC method has been routinely employed for testing new statistical models for data-retention of Flash devices [5, 6]. However, the accuracy of the EC technique as a reliability monitor for NVMs has not been demonstrated so far.

In this paper we show for the first time that characterization based on the EC concept yields the same statistical results as the full bit-by-bit analysis, thus validating the EC method as a fast yet accurate tool for memory characterization. Also, the EC method is applied for studying the reliability of ultra-scaled EEPROM devices. Finally, by using a statistical model for SILC, we show how the oxide degradation can be estimated from the SILC distribution obtained by the EC technique.

2. The EC approach

The gate-stress evolution of a large NVM array typically displays tail effects, which are determined by the statistical distribution of SILC in the array [7]. From the analysis of the gate-stress behavior of the array (particularly in the tail region) the SILC distribution can be obtained. For the characterization of SILC in a large number (typically of the order of 10^6) of bits, all $V_T - t$ characteristics of cells must be collected and manipulated for $I-V$ extraction [8]. Obviously, this may represent a very time- and memory-consuming task for most computational systems. To avoid numerical manipulation of large amounts of data, the EC technique for leakage characterization in

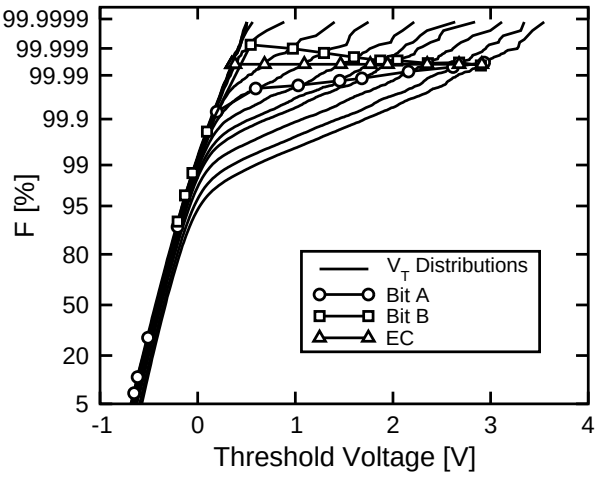

Figure 1. V_T distribution for a 768 kbit Flash array, with real cell paths (bits A and B) as compared to the EC approach.

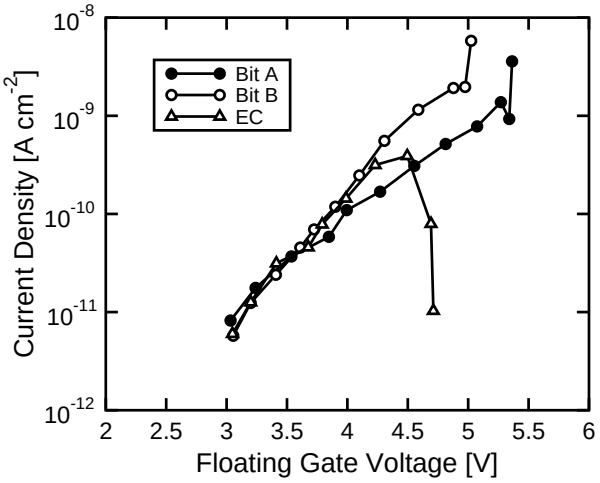

Figure 2. $I-V$ characteristics of the real bits A, B and of the EC dispayed in Fig. 1.

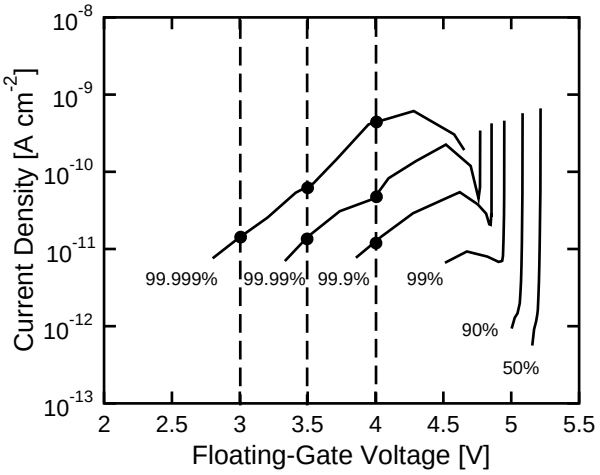

Figure 3. Selected ECs for a Flash EEPROM array with $t_{ox} = 8$ nm after 10^4 P/E cycles. SILC distribution is obtained by collecting the current density at fixed V_{FG}, as indicated by the vertical straight line.

Flash arrays was developed. According to this technique, $V_T - t$ characteristics are taken directly from the time-dependent distributions of V_T at a given cumulative probability F [4]. The fundamental assumption of this approach is that bits ordered by their V_T always maintain the same position within the cumulative distribution, irrespective of the gate stress time. The resulting $V_T - t$ are then converted into $I - V$s via the standard formulas [3]. Clearly, $I - V$ characteristics obtained by the EC method do not correspond to real bits in the array, rather they represent the average behavior of bits around a given F (hence the name *equivalent*).

2.1. Limitations for $I - V$ characterization

We analyzed some representative paths of tail bits in the V_T distributions during gate stress, in order to evaluate the impact of the constant-F approximation on resulting $I-V$ characteristics. Fig. 1 shows the V_T distributions for increasing gate-stress times, for a 768 kbit array of Flash EEPROM with oxide thickness $t_{ox} = 8$ nm after 10^4 P/E cycles [9]. Two bits (A and B) were selected in the tail, for which the position within the distribution at increasing times is shown as compared to a typical EC path. Clearly, the probability F for bits A and B is not conserved during gate-stress, indicating that the EC method can represent a poor approximation for some tail bit.

Fig. 2 shows the $I - V$ characteristics of real bits A, B and of the EC shown in Fig. 1. The EC approach largely underestimates the SILC above about 4.5 V of the floating-gate voltage V_{FG}. This can be understood by considering the bit paths in Fig. 1: tail bits A and B both start in the main part of the distribution, at relatively small V_T, where they also spend the early gate-stress times before joining the tail. On the other hand, the EC approximation starts at already large V_T maintaining contant F. Upon entering the tail, the EC is forced to an abrupt transition from very slow V_T drift (early times) to SILC (large times, full tail). This results in the large current increase in the EC characteristic at $V_{FG} \approx 4.5$ V. Therefore, only the EC $I - V$ below $V_{FG} = 4.5$ V is physically meaningful, whereas the higher V_{FG} range has to be discarded. Note also that raising the gate bias during gate stress results in an increase of the V_{FG} range for which the EC method works accurately. For instance, applying $V_{CG} = 8$ V would shift the maximum of the EC $I - V$ in Fig. 2 from 4.5 V to about 5.8 V, thus increasing the voltage range of accuracy of the EC characterization. Thus, the EC technique can provide a detailed description of the SILC distribution, over the whole voltage range of interest.

3. Accuracy of statistical characterization

For a statistical representation, any $I - V$ of real-bits/ECs must be reduced to a single parameter, such

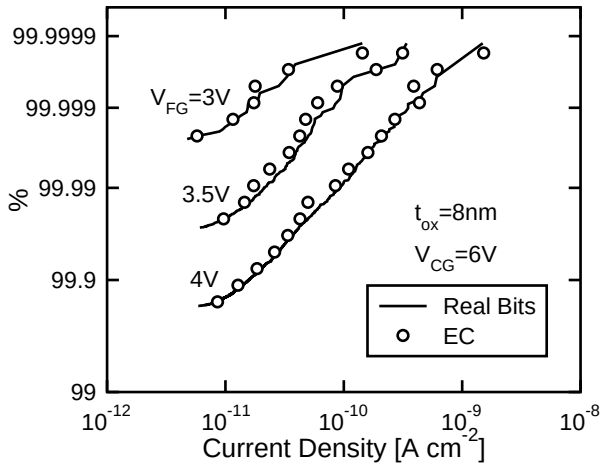

Figure 4. SILC distribution for ECs and real bits, for a Flash array with $t_{ox} = 8$ nm after 10^4 cycles. Distributions were obtained for different V_{FG}'s from gate-stress data under $V_{CG} = 6$ V.

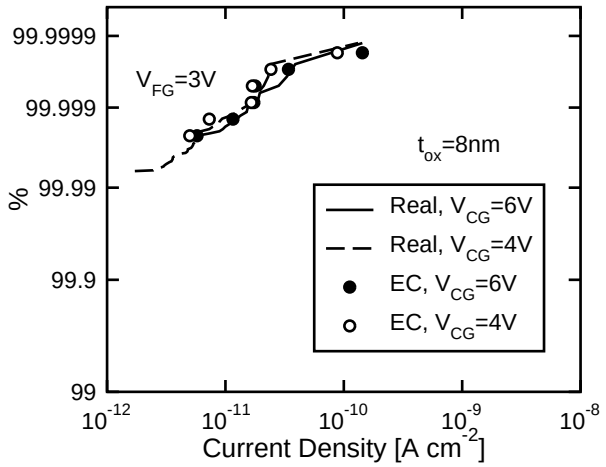

Figure 5. SILC distribution for ECs and real bits, for a Flash array with $t_{ox} = 8$ nm after 10^4 cycles. Distributions were obtained for $V_{FG} = 3$ V from gate-stress data under different V_{CG}'s.

as the FN parameter B related to the effective barrier height [7, 8], or the leakage evaluated at a given V_{FG}, which is generally used for reliability estimation and modeling [5, 6]. The latter distribution can be easily obtained collecting the SILC at fixed voltage, as shown in Fig. 3 for selected ECs with $t_{ox} = 8$ nm after 10^4 P/E cycles. Note that the extraction of SILC values has to be carried out at relatively small V_{FG}'s, where the EC $I - V$s provide a good approximation of average real-bits (see Fig. 2).

We compared the SILC distributions for ECs and real cells for a Flash array with $t_{ox} = 8$ nm after 10^4 P/E cycles. Fig. 4 shows SILC distributions for ECs and real bits, obtained from a gate-stress experiment under a control-gate bias $V_{CG} = 6$ V. A close agreement is found for all values of V_{FG} (3, 3.5 and 4 V corresponding to the vertical lines in Fig. 3). To demonstrate that the accuracy of the

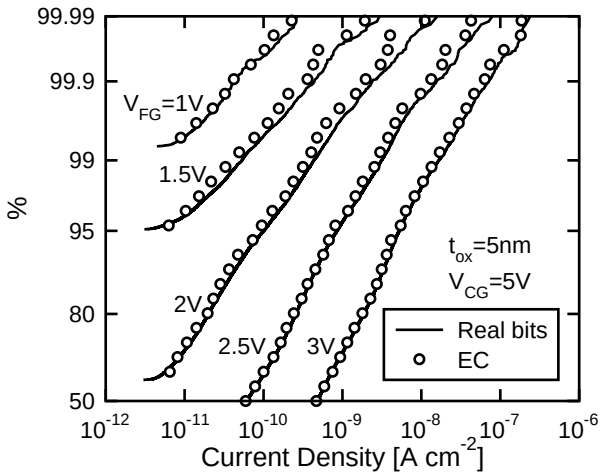

Figure 6. SILC distribution for ECs and real bits, for a Flash array with $t_{ox} = 5$ nm after 4×10^3 cycles and for different V_{FG}'s.

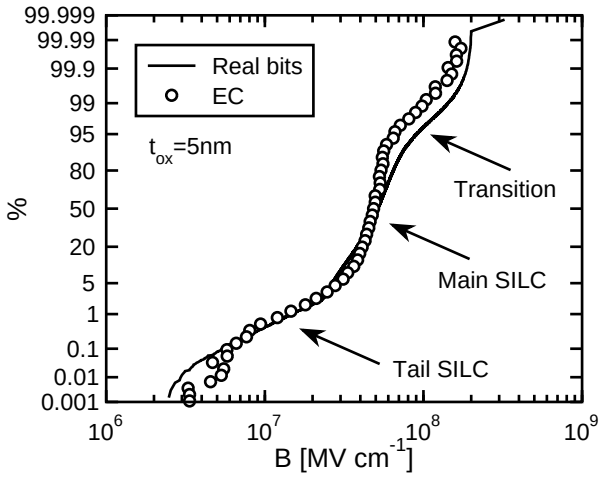

Figure 7. Distribution of B for ECs and real bits, for a Flash array with $t_{ox} = 5$ nm after 4×10^3 cycles.

EC method does not depend on experimental conditions, we compare in Fig. 5 SILC distributions for ECs and real cells, extracted from gate-stress data at $V_{CG} = 4$ and 6 V. Again, all distributions fall on the same line, supporting the accuracy of the EC technique for characterizing the SILC statistics in large arrays.

4. SILC in ultrascaled arrays

We applied the EC technique for studying the dependence of SILC on device scaling, by characterizing a 64 kbit Flash array with ultrathin tunnel oxide. Using a self-consistent Schrödinger-Poisson model for calculating the capacitance [10], we determined the oxide thickness to be $t_{ox} = 5$ nm. After 4×10^3 P/E cycling, we applied gate-stress conditions with $V_{CG} = 5$ V to the Flash array. Fig. 6 shows SILC distributions for ECs and resulting from a bit-by-bit approach, again demonstrating the high accuracy of the EC technique for all values of V_{FG} (1, 1.5,

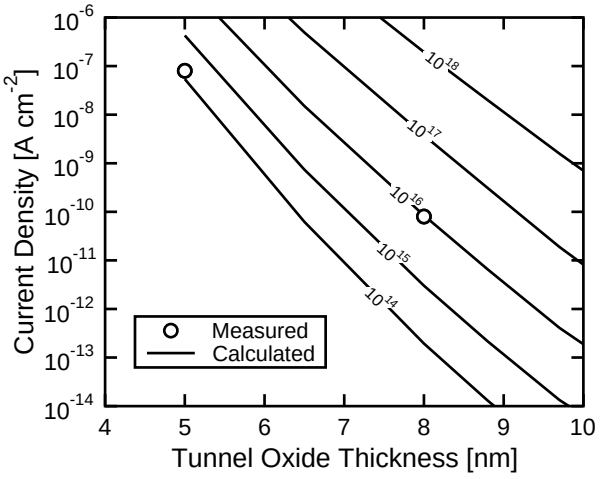

Figure 8. Calculated density of leakage current at oxide field $E_{ox} = 5$ MV cm^{-1} for a cumulative probability of 99.99%, as a function of t_{ox} and for increasing defect concentration. Also shown are experimental data for $t_{ox} = 5$ and 8 nm.

2, 2.5 and 3 V). As a further demonstration, Fig. 7 shows the distribution of the parameter B, *i.e.* the slope of the $JE_{ox}^{-2} - E_{ox}^{-1}$ characteristics (FN plot) [7]. A good agreement is obtained between the distributions of B for EC characteristics and for real cells. Note that both distributions display main and tail SILC regions, highlighting the ability of the EC technique to probe the leakage behavior within the array [7].

As demonstrated in Fig. 6, tunnel-oxide downscaling in Flash memories leads to a large enhancement of SILC. One reason for this severe increase of leakage is the strong dependence of tunneling lifetimes on the thickness of potential barriers [11]. However, the increased SILC tails may also be due to a larger defect generation in the thin oxide. To compare the oxide degradation in Flash arrays with different oxide thickness, we adopted a statistical model for SILC, which accounts for tunneling assisted by oxide defects and clusters of defects [5]. In the model, defect generation is assumed uniformely distributed in the tunnel oxide, and cluster formation is computed according to Poisson statistics. SILC is calculated based on accurate physical models for trap-assisted tunneling (TAT) and tunneling at multiple defects. This model was shown to accurately reproduce the statistical distribution of SILC in Flash arrays for oxide thickness in the range $6.5 - 9.7$ nm [5].

We calculated SILC distributions for oxide thicknesses 5 to 10 nm and defect concentration in the range $10^{14} - 10^{18}$ cm^{-3}, at an oxide field $E_{ox} = 5$ MV cm^{-1}. Fig. 8 shows the current density corresponding to a cumulative probability $F = 99.99\%$, as a function of oxide thickness. The figure also shows the current density measured at the same probability F in Figs. 4 and 6. These experimental data match the SILC calculation for defect densities of $10^{14} - 10^{16}$ cm^{-3}, with a smaller concentration for the 5 nm oxide. The lower P/E cycling (4×10^3 cycles) of the scaled device might partially explain the reduced degradation. However, a general decrease of defect density with decreasing oxide thickness was already observed in Flash memories at equal P/E cycling [5]. The results in Fig. 8 show how the EC characterization, together with a physical/statistical model accounting for the thickness dependence, can provide a valuable tool for estimating the oxide degradation in cycled memories.

5. Conclusions

In this paper the accuracy of the EC method has been directly demonstrated, by comparing the characterization results of the EC approach with the real distribution. Both the distribution of leakage currents and of the parameter B obtained by the EC method closely match the real distributions in the array. Finally we showed how the EC technique, in combination with a statistical model for SILC accounting for the SILC dependence on oxide thickness, can provide a straightforward tool for evaluating the oxide degradation in Flash memories.

6. Acknowledgements

The authors would like to thank Guoqiao Tao and Hein Valk for supplying the experimental data for the 8 nm samples.

7. References

[1] K. Naruke, S. Taguchi, and M. Wada, in *IEDM Tech. Dig.*, pp. 424–427, 1988.

[2] S. Satoh, G. Hemink, K. Hatakeyama, and S. Aritome, *IEEE Trans. Electron Devices*, vol. 45, pp. 482–486, Feb. 1998.

[3] A. Modelli, F. Gilardoni, D. Ielmini, and A. S. Spinelli, in *Proc. IRPS*, pp. 61–66, 2001.

[4] D. Ielmini, A. S. Spinelli, A. L. Lacaita, and A. Modelli, *IEEE Electron Device Lett.*, vol. 23, pp. 40–42, Jan. 2002.

[5] D. Ielmini, A. S. Spinelli, A. L. Lacaita, and A. Modelli, in *IEDM Tech. Dig.*, pp. 703–706, 2001.

[6] R. Degraeve, F. Schuler, M. Lorenzini, D. Wellekens, P. Hendrickx, J. V. Houdt, L. Haspeslagh, G. Groeseneken, and G. Tempel, in *IEDM Tech. Dig.*, pp. 699–702, 2001.

[7] D. Ielmini, A. S. Spinelli, A. L. Lacaita, L. Confalonieri, and A. Visconti, in *Proc. IRPS*, pp. 73–80, 2001.

[8] N. Tsuji, K. Sakakibara, N. Ajika, and H. Miyoshi, in *1998 Symp. VLSI Tech. Dig.*, pp. 196–197, 1998.

[9] G. Tao, H. Valk, A. Scarpa, L. van Marwijk, F. Kuper, G. Dormans, and R. Verhaar, in *Proc. Non-Volatile Semiconductor Memory Workshop*, pp. 130–132, 2001.

[10] A. S. Spinelli, A. Pacelli, and A. L. Lacaita, *IEEE Trans. Electron Devices*, vol. 47, pp. 2366–2371, Dec. 2000.

[11] I. Lundström and C. Svensson, *J. Appl. Phys.*, vol. 43, pp. 5045–5047, 1972.