Semiconductor Physics, Quantum Electronics & Optoelectronics. 2004. V. 7, N 1. P. 112-117.

PACS: 85.40.-e

Fabrication and characterization of ONO and tunnel oxide for 16k FLOTOX EEPROM cell

U. Hashim[1], R.M. Ayub[2] and K.S. On[3]

[1]*School of Microelectronics Engineering, Northern Malaysia University College of Engineering, Kubang Gajah 02600 Arau, Perlis, Malaysia, E-mail: uda@kukum.edu.my*
[2]*NVM Research Group, MIMOS Semiconductor, MIMOS Berhad, Technology Park Malaysia, 57000 Kuala Lumpur, Malaysia*
[3]*Wafer Process Technology Group, Malaysia Microelectronic Solution Sdn Bhd Suite S01, 2nd Floor, 2300 Century Square, 63000 Cyberjaya, Selangor, Malaysia*

Abstract. The EEPROM process is one the hardest process to be developed. The performance of the EEPROM devices is normally judged on the programming speed, which relates to program high (erase) and program low (write) operations. It is essential that the program high and program low speed of the EEPROM cell is within 1ms with a programming voltage of not more than 16V. In this study, two experiments were setup to improve the programming speed. The first experiment was to increase the high voltage NMOS drain junction breakdown voltage with the source floating (HVNMOS BVDSF), and the second experiment was to scale down the ONO layer. The characterization work to increase the programming speed of the memory cell of 16k FLOTOX EEPROM has been carried out. *P*-field implant dose is optimized to have both the HVNMOS BVDSF and the p-field threshold voltage above 16V for fast programming. As a result, the threshold voltages of programming high and low operation are achieved at 4.35V and –0.77V respectively. Furthermore, by scaling down the nitride layer of ONO from 160Å to 130Å, the V_t program window is further improved to 4.5V and –0.94V for the program high and program low operations respectively.

Keywords: EEPROM, memory cell, threshold voltage, program high, program low, ONO, nitride layer.

Paper received 01.08.03; accepted for publication 30.03.04.

1. Introduction

EEPROM is classified as a non-volatile memory product; it refers to Electrically Erasable and Programmable Read Only Memories, and is available in two types, ie embedded and stand-alone [1]. Compared to a CMOS process, which only has 2 devices i.e. NMOS and PMOS, an EEPROM chip, is constructed from 6 devices, namely the low voltage NMOS (LVNMOS), the low voltage PMOS (LVPMOS), the high voltage NMOS (HVNMOS), the high voltage PMOS (HVPMOS), the natural NMOS (ZMOS), and the memory cell (EPROM cell).

The performance of the NVM products is normally judged from the programming speed and the density of the memory [2, 3]. Besides conventional device characteristics, nonvolatile memory cells also have additional functional memory characteristics, which are used to evaluate the performance of the memory cell. One of the most important parameters is programming speed versus programming voltage, which describes the time dependence of the threshold voltage with respect to programming time, at different programming voltages.

In the EEPROM product under study, an internal timer limits the programming time to 1 ms, so a large amount of carriers need to be pumped in or out of the floating gate in a relatively short period of time. The threshold voltage of the EEPROM cell transistor will change accordingly. The efficiency of the pumping of carriers in and out of the floating gate is described as the programming speed.

A typical memory cell layout, called the Floating Gate Tunnel Oxide cell (FLOTOX) is shown in Fig. 1 and the cross-section of the memory cell device is illustrated in Fig. 2. The memory cell device consists of a stacked gate MOS structure with a small thin tunnel oxide region in the floating gate/drain overlap area. The

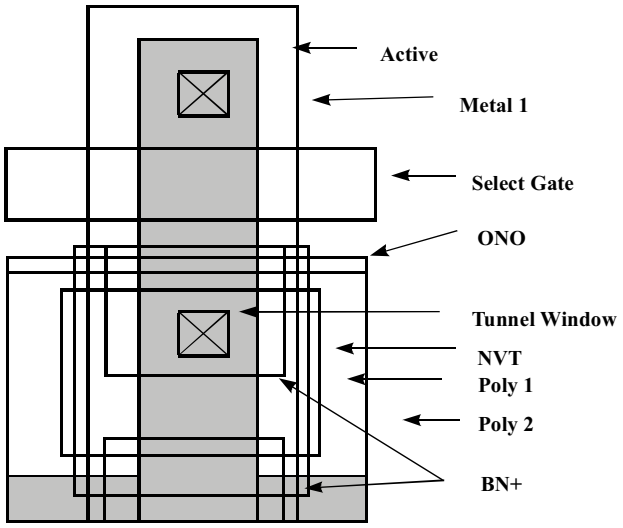

Fig. 1. Typical of memory cell layout.

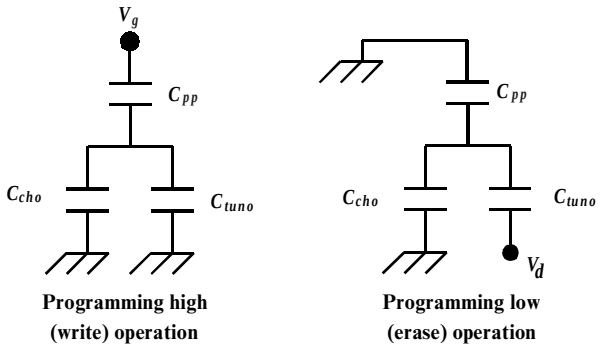

Fig. 3. Simplified equivalent circuit of the FLOTOX EEPROM during the program high and program low operations.

lower polysilicon gate (the floating gate) is completely surrounded by silicon dioxide, whose quality and integrity enables charge to be retained in the floating gate for more than 10 years. The programming voltage is applied between the control gate and the drain, and is capacitively coupled to the tunnel oxide. The cell is programmed high (logic 1) by causing electrons to tunnel into the floating gate from the drain by applying a positive voltage on the control gate with the drain grounded and the source floated. The cell is programmed low (logic 0) by causing electrons to tunnel from the floating gate to the drain by applying a positive voltage to the drain with the control gate grounded and the source floated. The number of electrons trapped in the floating gate determines the threshold voltage of the EEPROM cell, which is detected in the read operation with suitable voltage acting on the control gate [4, 5].

Fig. 3 shows the simplified equivalent circuit of the FLOTOX EEPROM transistor during program high and program low operations respectively. Cpp is the poly to poly overlap capacitance, Ctuno is the tunnel window capacitance and Ccho is the channel capacitance [6].

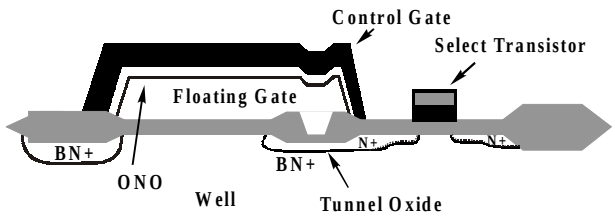

Fig. 2. Cross-section of memory cell of EEPROM device.

The fraction of the applied programming voltage falling across the tunnel oxide can be calculated according to the fixed coupling ratios, KH during the program high operation and KL during the program low operation which are derived from the oxide capacitances (please also refer to Fig. 2) [5,6,7] as follows:

$$K_H = \frac{C_{pp}}{C_{total}}, \tag{1}$$

$$K_L = \frac{C_{pp} + C_{cho}}{C_{total}} \tag{2}$$

and $$C_{total} = C_{pp} + C_{cho} + C_{tuno} \tag{3}$$

where K_H = Coupling ratio during program high, K_L = = Coupling Ratio during program low

During the program high operation, the voltage on the tunnel oxide, VtunH is given by:

$$V_{tunH} = K_H V_g + \frac{Q_{fgH}}{C_{total}}, \tag{4}$$

where Q_{fgH} = floating gate charge, V_g = applied gate voltage.

For a programming pulse of width t, the stored charge in the floating gate shifts the program high and program low threshold voltage denoted as V_{tH} and V_{tL} respectively as follows [5,6,7].

$$V_{tH} = V_{ti} + V_g - \frac{\beta X_{tun}}{K_H \ln\left\{\left(\frac{A_{tun}\alpha\beta}{X_{tun}C_{totalo}}\right)t + \exp\left(\frac{\beta X_{tun}}{K_H\left(V_g - V_{ti} + V_t H(0)\right)}\right)\right\}} \tag{5}$$

$$V_{tL} = V_{ti} + V_d \frac{K_L}{K_H} - \frac{\beta X_{tun}}{K_H \ln\left\{\left(\frac{A_{tun}\alpha\beta}{X_{tun}C_{total}}\right)t + \exp\left(\frac{\beta Xtun}{K_H\left(V_L K_L - K_H V_{ti} + K_H V_{tL}(0)\right)}\right)\right\}}, \quad (6)$$

where X_{tun} = tunnel oxide thickness α and β are the characteristic constants of the tunnel oxide in the Fowler Nordheim tunnelling equation, V_{ti} = threshold voltage with zero charge in the floating gate, A_{tun} = tunnel area, E = cathode field through tunnel dielectric, $V_{tH}(0)$ = threshold voltage at zero time assuming the cell has already been programmed low, $V_{tL}(0)$ = threshold voltage at zero time assuming the cell has already been programmed high

2. FLOTOX EEPROM cell programming speed

In principle, the programming speed of the EEPROM cell depends on tunnel Oxide Thickness (X_{tun}), programming voltage (V_p), ONO thickness (X_{pp}) and Poly to Poly overlap Area (A_{pp}).

The thinner the tunnel oxide is, the faster the programming speed. However, due to the charge retention requirement, the tunnel oxide thickness cannot be reduced beyond 85Å. Therefore 85Å is the nominal thickness throughout the experiment and will not be changed.

The poly to poly overlap area is fixed in the design and will not be experimentally changed.

The programming speed depends critically on V_p. At V_p = 16V, it takes about 1ms for V_{tH} to reach 4.5V, whereas at V_p = 13V it takes more than 1 sec. In a conventional EEPROM circuit, the charge pump is designed to pump V_p to above 20V. The actual V_p is, however, clamped by the reverse n+ p-junction of the drain side of the HVNMOS. HVNMOS appears in the charge pump circuit as well as in the decoder logic and in the selection transistors.

The clamp voltage depends on the reverse breakdown voltage of the n+p-diode of the drain side of the HVNMOS (HVNMOS BVDSF, where BVDSF refers to the drain junction breakdown with source floating), which depends mainly on the p-(substrate) dopant concentration surrounding the n+junction. The higher the substrate dopant concentration is, the lower the breakdown voltage is.

The region of the n+ junction in conjunction with the field oxide has the lowest breakdown voltage. In order to raise the field threshold voltage, the boron concentration of the p-field area is raised by a p-field implant step. If the concentration is too low, the p-field threshold will be less than V_p. This will cause the parasitic MOS in the field region between two n+ diffusion regions to be turned on.

Here comes a dilemma. The higher the p-field implant dose is, the higher the p-field V_t is, but the lower the HVNMOS BVDSF is. In order to have both the HVNMOS BVDSF and the HV p-field threshold to be above 16V, an experiment to trim the p-field implant dose and the field oxide thickness has to be done.

3. Experiment methodology

3.1. Fabrication of the FLOTOX EEPROM cell

The EEPROM memory cells under study are fabricated on p-type silicon wafers <100> with the resistivities of 16–25 Ohm·cm. The fabrication of the memory cell started with implantation of arsenic to form the buried N+ region. Thermal oxide was subsequently blanket grown to a 600Å thickness on the buried N+ region. The tunnel window was then patterned where ultra thin tunnel oxide was immediately grown to a thickness of 85Å. The formation of the floating gate was then carried out by the deposition of POCL3 doped polysilicon with a thickness of 2300Å. ONO was used for inter-poly dielectric to insulate the floating gate from other electrodes. The ONO dielectric consists of a 100Å thermal oxide layer grown at 980°C in an O_2/N_2 environment, an LPCVD nitride layer of 130Å thickness deposited on top by CVD furnace and lastly a thermal oxide layer of 20Å thickness. The inter-poly dielectric construction was accomplished with the patterning of ONO. Next the formation of the control gate was done by deposition of polysilicon of 4000? thickness using LPCVD. The fabrication process was continued with conventional backend processes for the interconnect construction to complete the whole circuitry. The n-channel memory cell transistor structures are now ready to be optimized for programming speed improvement.

3.2. Characterization of the FLOTOX EEPROM cell

The setup for the measurement of programming speed is as follows. An Agilent 4155C DC Parametric Analyser was used to generate the "high" and "low" waveform signals for the programming operation. The same equipment was used to measure the cell threshold voltages. First the EEPROM cells were exposed to a 5 ms program low pulse of 16V to bring the V_t to a low level. The cell was then programmed high with a fixed programming voltage but with the variable programming pulse width. The programming pulses increased so that the cumulative programming time falls in logarithmic increments of 0.5 ms, 1 ms, 2 ms, 5 ms, 10 ms …100 ms. Vt was measured at each of these cumulative intervals. After a com-

plete sweep of exposure times up to 100 ms, the programming voltage was incremented and the process was repeated again. Similar procedures were used to measure the program down speed by programming the cell high at the start and then programming low successively.

The second experiment was to change the nitride thickness in the ONO layer and to measure the programming speed with different nitride thickness.

4. Results and discussions

Fig. 4 is the focus ion beam cross-sectional view of the FLOTOX EEPROM cell. This figure clearly shows a complete FLOTOX EEPROM cell, which consists of a control gate, a floating gate, a select gate, a tunnel window, an ONO layer, a source area and a drain area. The EEPROM cell is sitting on a buried N+ layer. The second part of Fig. 4 is the corresponding Supreme simulation. It can be seen in the 2D Supreme simulation graph that the highest *p*-dopant concentration surrounding the drain side of the selection transistor is in the field oxide boundary, the concentration reaches about $2 \cdot 10^{17}/cm^3$ with a *p*-field implant dose of $7.5 \cdot 10^{13}/cm^2$.

Fig. 5 and Fig. 6 show the relationship of the HV-*p*-field implant dose, to the HVNMOS BVDSF and HV *p*-field V_t. It can be seen clearly that the higher the dose is, the higher the HV *p*-field V_t is but the lower the HVNMOS BVDSF is. Fig. 5 indicates that for field oxide of 5500Å, the two lines meet at a point of 15.6 V, which is less than the target of 16V. Since 5500Å thick field oxide did not meet the V_t target, a 7000Å field oxide step was developed. Fig. 6 shows that with 7000 Å, both the HVNMOS BVDSF and HV-*p*-field Vt are above 16V, being 17.3V and 22.4V respectively, at an implant dose of $6 \cdot 10^{13}/cm^2$.

Fig. 7 is the programming high speed of the memory cell transistor. It shows that at 16V programming voltage and 1ms programming time, a high threshold voltage high (V_t–high) of 4.2V is achieved. The V_t at the 0.01 ms programming time is the V_t after UV light exposure. The kinks at 0.5 ms programming time are due to equipment limitations whereas 0.5 ms is the minimum programming time.

Fig. 8 is the programming low speed of the same cell transistor, which was targeted to be of negative threshold voltage in 1 ms programming time. It shows that at 16V programming voltage and 1 ms programming time, the V_t–low is –1.5V.

The next step to improve the programming performance was to scale down the ONO layer. Only the nitride layer was scaled down in the experiment. Before scaling down, the original nitride thickness in the ONO layer was 230Å. After scaling down, the final optimized thickness of the ONO layer must be able to withstand 16V for reliability purpose. Fig. 9 illustrates the effect of different nitride thickness in the ONO layer to V_t–high, Vt–low and Fig. 10 illustrates the effect of different nitride thickness in the ONO layer to the ONO breakdown voltages.

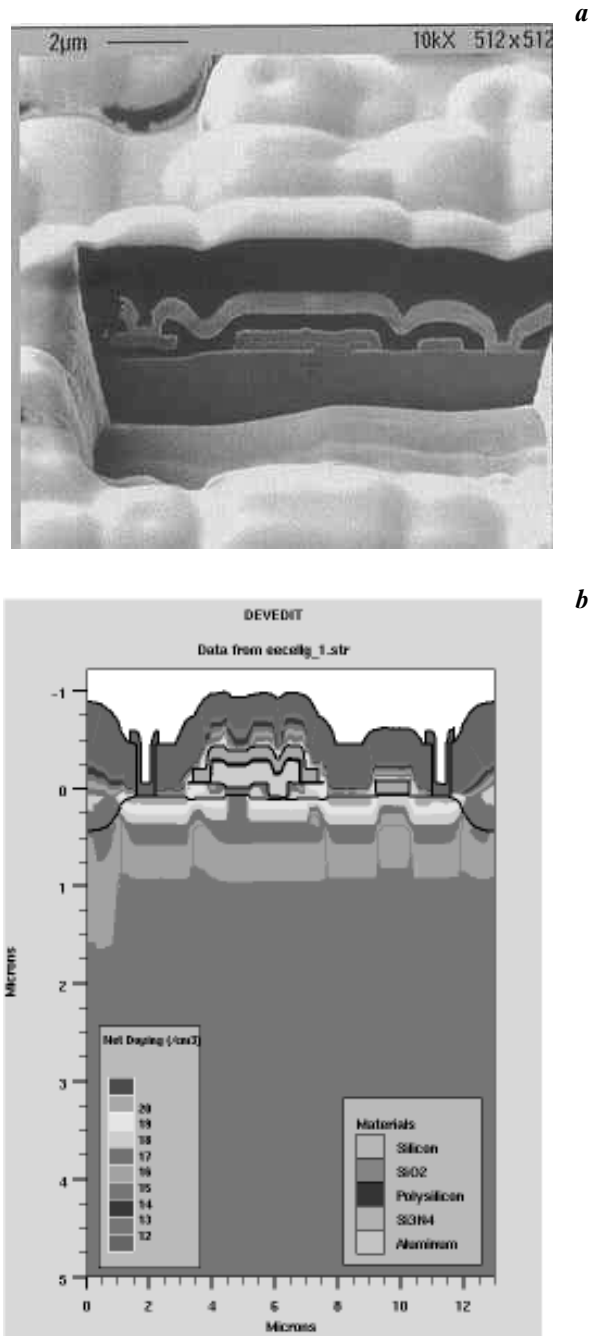

Fig. 4. FIB cross-sectional view of a FLOTOX EEPROM cell (***a***) and the Supreme simulation of the same cell (***b***).

The experiment on the programming speed versus ONO thickness shows that thinner ONO gives faster programming speed. However the specification of ONO breakdown voltage is set at 16V. In such a case, 130Å is chosen to be the standard nitride thickness. With this thickness V_{tH} reaches 4.6V and V_{tL} reaches –0.9V in 1ms. This is enough for successful EEPROM operation.

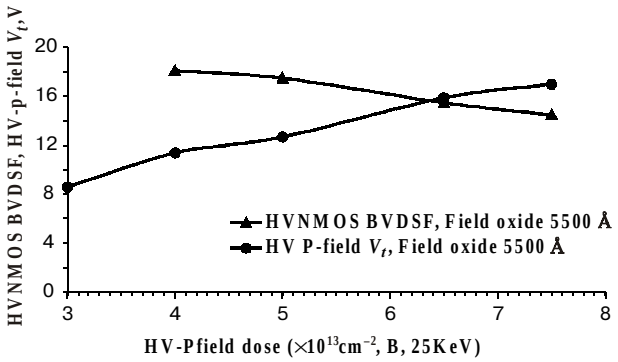

Fig. 5. The relationship of HVNMOS BVDSF and HV-Pfield V_t versus HV-Pfield implant dose for field oxide thickness of 5500Å.

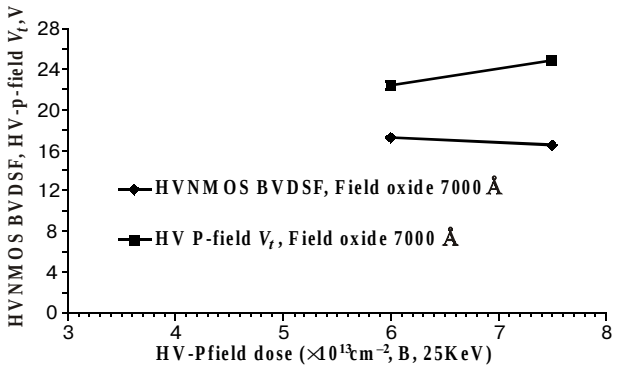

Fig. 6. The relationship of HVNMOS BVDSF and HV-Pfield V_t versus HV-Pfield implant dose for field oxide thickness of 7000Å.

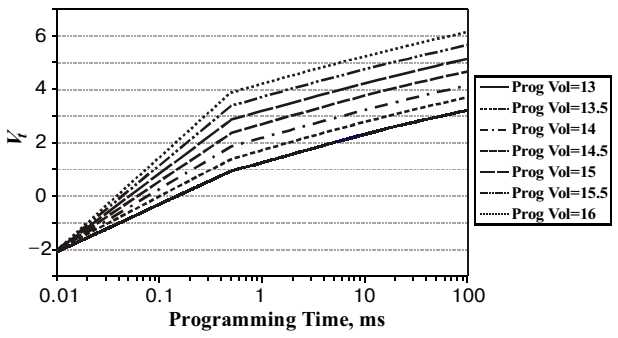

Fig. 7. Measured threshold voltage shift as a function of program up time for different programming voltages.

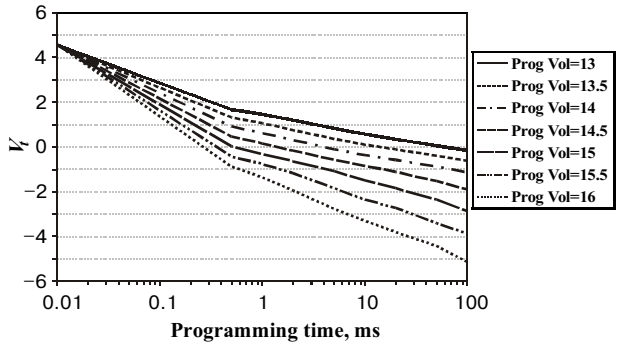

Fig. 8. Measured threshold voltage shift as a function of program low time for different programming voltages.

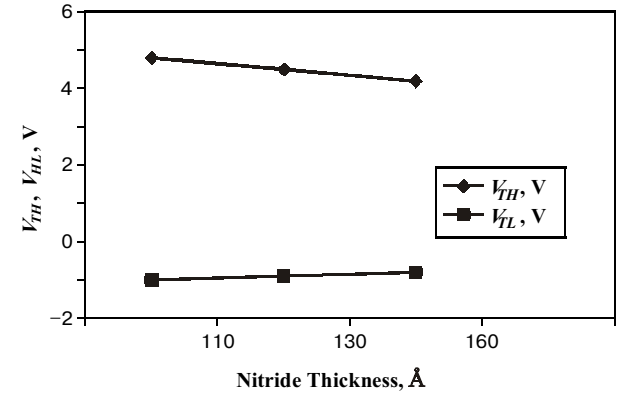

Fig. 9. The high and low cell threshold voltage versus the nitride thickness in the ONO layer.

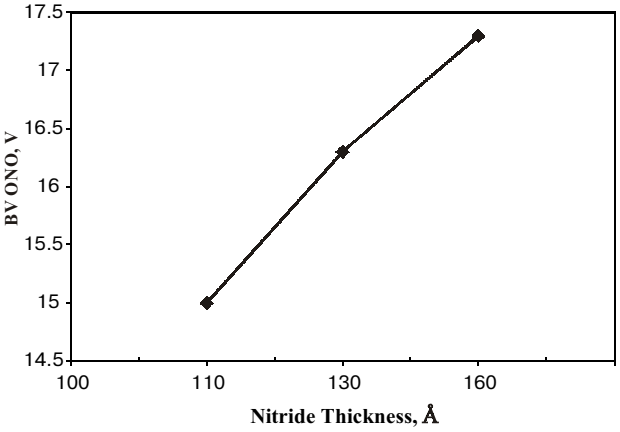

Fig. 10. ONO breakdown voltage versus nitride thickness in the ONO layer.

5. Conclusions

The characterization work to increase the programming speed of the memory cell of 16k FLOTOX EEPROM has been carried out. Boron concentration of $6{\cdot}10^{13}$ for *p*-field implant was selected to have both the HVNMOS BVDSF and *p*-field V_t above 16V. As a result, the threshold voltages of programming high and low operation are achieved at 4.35 V and –0.77V respectively. Furthermore, by scaling down the nitride layer of ONO from 160 Å to 130 Å, the Vt program window is further improved to 4.5V and –0.94 V for the program high and program low operations respectively.

Acknowledgement

The authors would like to thank MIMOS Berhad FAB1, Failure Analysis lab and Test lab personnel for their technical advice and contributions.

Reference

1. William D. Brown and Joe E. Brewer, Nonvolatile semiconductor memory technology, a comprehensive guide to understanding and using NVSM devices // *IEEE Press*, pp. 1-149.
2. P. Canet, R. Bouchakour and J.M. Mirabel, EEPROM programming study – time and degradation aspects // *IEEE Letter*, 2001, pp. 846-849.
3. Robert Versari, Augusto Pieracci and Bruno Ricco, Fast programming/erasing of thin-oxide EEPROMs // *IEEE Transactions on Electron Devices*, vol. **48**, No. 4, April 2001, pp. 817-820.
4. Dennis Walker, Jr. Brad Skotko and Steven Bibyk, Charactericzation of two standard CMOS EEPROM designs // *IEEE letter*, 2001, pp. 894-896.
5. A. Bhattacharyya, Modelling of write/erase and charge retention characteristics of floating gate EEPROM devices // *Solid State Electrons*, vol. **27**, No.10 pp 899-906, 1984.
6. A. Kolodny, S.T.K. Neih and B. Eithan, Analysis and modeling of floating gate EEPROM cell // *IEEE Trans. Electron Device*, vol. **ED33** No. 6, pp. 835-844, June 1986.
7. G. Yaron, S.J. Prasad, M.S. Ebel and B.M.K. Leong, A 16k EEPROM employing new array architecture and design-in reliability features // *IEEE J. Solid State Circuits*, vol. **SC-17**, No. 5, pp. 833-840, 1982.