Computer and experimental study of the gate dielectric in a memory transistor

R. Avichail-Bibi[1], D. Fuks[1], A. Kiv[1] & Ya. Roizin[2]
[1]*Ben-Gurion University of the Negev, Israel*
[2]*Tower Semiconductor Ltd., Israel*

Abstract

We demonstrate a novel approach that enables combining microscopic studies of the behaviour of the injected charge (IC) in the gate dielectric (GD) of the memory transistor and the description of kinetics of memory device service parameters. To study the microscopic processes of the redistribution of the IC in the GD a special package of programs was developed that allows the modelling the migration of injected electrons and holes in the GD. The model accounts real properties of dielectric (spatial distribution of local centres and their characteristics, the dielectric constant and its changes on the microscopic distances, a complex composition of dielectric, temperature conditions and the geometry of the GD). The results of the computer simulation of microscopic characteristics of the IC were used as input data for the commercial Device simulation program "Medici". We found a correlation between microscopic characteristics of IC in GD and the service parameters of the memory device and realized the feedback procedure changing the GD characteristics in the simulation model.
Keywords: memory transistor, gate dielectric, trapping mechanisms, molecular dynamics, computer simulation.

1 Introduction

The NROM (nitride read only memories) are non-volatile memories with local storage of charge at the edges of the memory transistor channel and a thick (>35 Å) bottom oxide (BOX) [1] that became popular in the nonvolatile semiconductor memory market. A ONO (SiO_2-Si_3N_4-SiO_2) stack with non-tunnel bottom oxide is the GD in a two-bit per cell memory transistor.


WIT Transactions on Modelling and Simulation, Vol 46, © 2007 WIT Press
www.witpress.com, ISSN 1743-355X (on-line)
doi:10.2495/CMEM070071

Information is stored as the charges injected into the Nitride at the channel edges of the memory transistor. Programming is performed by channel hot electrons. Holes, created by band-to-band tunneling in the drain region erase the programmed bit [1–3]. The device is read out in the "reverse" direction compared to programming.

These memories are free of the limited retention and high read disturb of the previous SONOS generations having thin (~20 Å) BOX. Properties of traps are especially important in NROM devices because the reliability performance of these memories is to a great extent determined by the lateral migration of the charge trapped in the silicon nitride layer [2]. Reliable values of trap parameters in silicon nitride, in particular the trap activation energy, can be found from device measurements, as well as the effects related to BOX can be distinguished from those connected with silicon nitride properties.

Despite numerous experimental data on traps collected for different types of silicon nitride, the chemical nature of traps in a Si_3N_4 remains unclear. This is why the phenomenological description and computer simulation of trapping processes in Si_3N_4 stack is of great interest. In [4] a new approach in Molecular Dynamics (MD) simulation is described. It allowed the study of real physical processes in the GD of memory device and the prediction of its retention-endurance characteristics.

Retention of *micro*FLASH® memory transistors is characterized by the stability of the programmed and erased state threshold voltages (V_t). The V_t shift of the programmed state of the device after cycling (a series of programming-erase (P/E) operations) is experimentally controlled after high temperature bakes. The V_t shift of the programmed state is sometimes called a high temperature, HT V_t shift.

The second type of V_t shift characterizes the erased state of the device after cycling. This shift has weak temperature dependence and is called room temperature, RT V_t shift. HT and RT V_t shifts determine the operation margins of the memory cell [5].

Analysis of time dependences of HT V_t shift (ΔV_t) shows two stages in its kinetics: "fast" V_t shift and "slow" V_t shift. The "fast" V_t shift depends on the programming window $\Delta V = V_{t\ high} - V_{t\ initial}$ and is typically 50 mV – 300 mV for ΔV=1-3 V. At a temperature ~200^0C the "fast" period of V_t decrease lasts several hours [14–16]. The "fast" relaxation period is followed by a "slow" V_t decrease process with high activation energy (~1.8eV) [6]. At the beginning of the relaxation process there is a limited decrease of V_t (~50-200 mV) even for the one-time programmed cell. Additional "fast" memory window loss (~250-500mV) is observed after 1k-100k program-erase cycles [5–7].

It was found in [8] that at least for a small ($<10^4$) number of cycles the HT V_t shift is dominated by lateral spread of the charge carriers trapped in the nitride layer of ONO. For computer modelling of such processes we developed software that is suitable for the investigation of the following processes and phenomena:

- Migration of the carriers in the dielectric with different types of potential relief;

- Thermalization of hot carriers;
- Influence of internal and external fields on the migration of carriers;
- Influence of geometric factors on the redistribution of injected carriers in the GD.
- Influence of physical characteristics of the processes in the memory device on its service parameters.

The adequate physical model and a simulation approach are described in [4, 8]. In this paper we present additional considerations of the mechanisms of retention loss by memory devices.

2 Spreading of injected charge in ONO induced by Coulomb repulsion

2.1 Non-Gauss distribution of injected carriers in GD

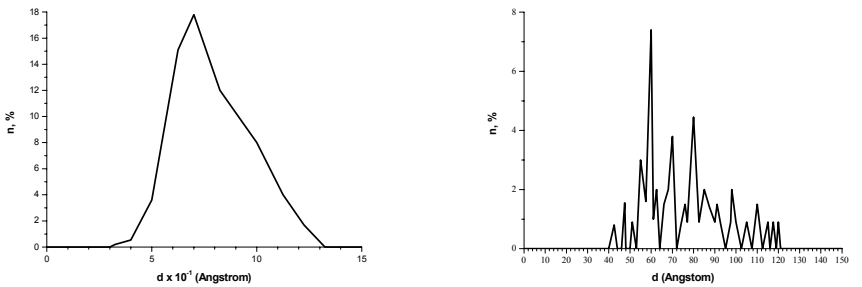
The original simulation program "Memory" [4, 8] was used to find real mechanisms of spreading of the IC in the GD. First, the diffusion profiles were studied for a wide range of the density of the IC and for a wide range of physical conditions in the GD. In these simulations the electrons or holes were injected from the bottom side into the horizontal ONO stack of ~150-200 Å length. The stack was divided into vertical slabs. The relative fraction of the IC with respect to their total number (n in %) was calculated in each slab that allowed construction of the distribution profile of IC. Dividing into thinner slabs gives possibility to visualize the fine structure of the distribution of the IC. We also varied the potential relief (PR) in dielectric, temperatures of carriers and of the ONO stack, types and spatial distribution of traps.

Making use of the scaling down procedure for the IC profiles we were able to follow the behaviour of the IC in the spatial intervals on an atomic scale. The results displayed in Figs. 1 and 2 show that the distribution of carries is discontinuous. In Fig. 1b we observe sub-peaks located at difference distances from each other. The next significant fact is that the distance between the sub-peaks does not depend on the temperature but depends on the density of the IC. The typical characteristic of IC profiles is the comparatively large charge droplets at some distances from the central peak. These charge clusters were named "parasitic peaks" (PP). Fig. 2 demonstrates the location of the profiles of the density distribution for the injected electrons and holes. Such a configuration of the density profiles for the injected charges corresponds to the erasing of the programmed state of the memory device. One can see in the right side of the distribution shown in Fig. 2 a separately located small sub-peak that corresponds to PP. Appearance of PP indicates clearly that the spreading of IC in the GD does not satisfy the Fick laws. We studied conditions of PP formation and found that they play a decisive role in mechanisms of retention loss by *micro*FLASH®

memory transistors and determine the double-programming effect in the memory device [9]. Formation of PP allows also the explanation of the existence of two limits of scaling down of memory devices with high-k layers in the GD [10]. Fig. 2 demonstrates the situation when the spatial location of electron and hole distributions do not overlap. This leads to continuous accumulation of carriers in the injection region of the GD in conditions of cycling work of the device.

The non-Gauss distribution of spreading IC decreases the efficiency of the erase processes because some PPs are located comparatively far form the IR. This is a source of the instability of the work of the device. Such a situation leads to the necessity to provide the optimal spatial and energy distribution of trapping and scattering centres in the GD and the optimal value of their density.

The relation between densities of trapping and scattering centres is of crucial importance due to the important role of scattering centres in the thermalization of hot IC.



a) b)

Figure 1: Two distribution profiles for injected electrons in the ONO stack of length 150 Å. The vertical slabs in ONO are of thickness a) 10 Å; b) 2 Å. d is the length in the ONO stack.

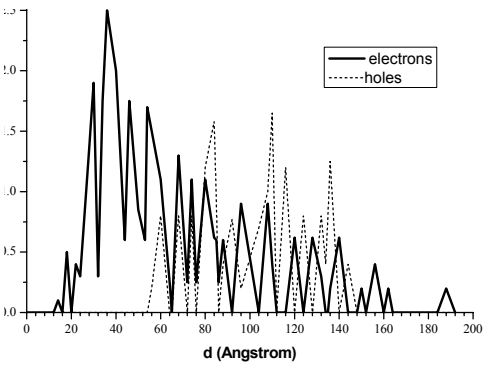


Figure 2: The distribution profiles for injected electrons and holes in the ONO stack of length 200 Å.

2.2 Kinetic features of spreading of injected carriers in ONO

The dependence of charges that leave the IR during some time interval on the value of injected charges was determined using our simulation program Memory. A number of charges were put up in the IR. Then the time intervals ($\Delta\tau_i$) between two nearest events of the IC leaving the IR were fixed. Simulation results show that the distribution of $\Delta\tau$ is described by an exponential law. This means that elementary acts of the escape of the IC from the IR occur according to exponential kinetics. This allows one to assume that the number of spreading electrons $\Delta Q \sim Q$, where Q is the number of injected electrons. Such a correlation determines the exponential dependence of Q on time, t.

Proceeding from the obtained kinetics we used the correlation $\Delta V_t \sim V_t$ in the mathematical description of the kinetics of V_t shift in [8]. These correlations are confirmed by the results of device modelling using the program Medici [11] (See Fig. 3). In Fig. 3 the voltage-current (V-I) characteristics are shown for different values of charge injected into the GD. Analysis of these graphs shows the relation $Q \sim V_t$. Consequently, ΔQ is proportional to ΔV_t. The inclination of V-I characteristics decreases from the left to the right. The inclination $dI/dV = 1/\rho$, where ρ is a specific resistance of the channel region of *n-p-n* transistor. Hence ρ increases in this direction. In this case the re-compensation of the channel region and the opening of the channel of the transistor demands larger V_t. Thus we see that from the one hand the simulation results obtained by our program Memory confirm the relation $\Delta Q \sim Q$ and from the other hand the Device simulation program Medici allows confirmation of the relation $Q \sim V_t$, and hence $\Delta Q \sim \Delta V_t$. These relations are used as a basis to derive the equation for the fast V_t shift in conditions of cycling [8, 12].

The results obtained by the program Medici are based on using the IC profiles obtained by the program Memory. Thus the non-Gauss profiles are accounted for in the device simulation by the program Medici. The device simulation results are consistent with experiment [13].

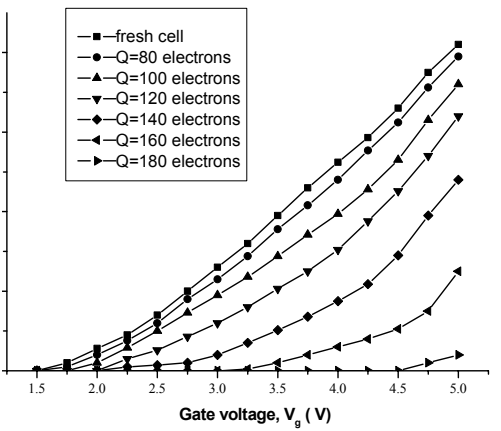


Figure 3: The voltage-current characteristics obtained with the program Medici for the fresh cell and for different numbers of IC.

3 Dependence of IC profiles on the features of potential relief in the GD

We performed simulation studies of the kinetics of spreading of the IC from the IR for different modifications of the potential relief (PR), in particular for different densities of traps in the GD. It was found that the activation energy for the carrier migration in the GD is an effective value and depends on the density of traps. This dependence becomes important for large densities of traps (more than 10^{20} traps/cm^3 [14]. This is one reason that the results obtained by different authors in some cases do not coincide. The simulation study of the dependence of the activation energy on the density of traps up to $5 \cdot 10^{21}$ traps/cm^3 was performed. In the region of some critical density of traps $N^*_{traps} \approx 10^{20}$ traps/cm^3 the PR sharply changes. The transformation of PR at $N_{traps} > N^*_{traps}$ was manifested by the significant increase of activation energy for carrier migration [15].

A strong dependence of the PR transformation on the degree of the overlapping of electron wave functions for the neighbor traps and on the value of potential wells of the individual traps was revealed. The overlapping of wave functions was modeled by changing the probability of tunneling between the nearest traps [4].

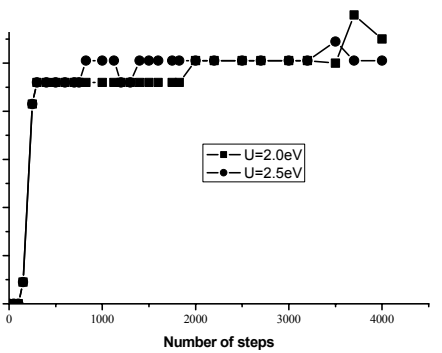


Figure 4: The kinetics of charge leaving the IR for the large number of traps: Ntraps > N*traps.

The increase of the depth of potential wells in PR of GD can be explained by formation of a strongly disordered PR as a result of the random overlapping of wave functions of neighboring trapped electrons. Thus we can suppose that in the case when $N_{traps} > N^*_{traps}$ the Anderson localization [16] occurs decreasing sharply the mobility of carriers. This effect leads to significant growth of the activation energy for carrier migration. Such a situation reveals itself in the fact that the kinetics of redistribution of the IC in the GD becomes independent on the activation energy of individual traps. Fig. 4 illustrates these results. One can

see that at the first steps of the simulation procedure the escape of some number of the IC from the IR occurs. After that we do not observe the redistribution of spreading IC. We see almost the same kinetics of the IC spreading for two different activation energies, U of individual traps (2 eV and 2.5 eV). These results show the way of the modification of the PR of the GD to achieve a stable distribution of the IC in the GD for a long time: it is necessary to create the disordered potential relief for electron migration with deep enough wells for Anderson localization.

The next example of the influence of the features of PR on the kinetics of the IC in the GD is linked to the creation of high-k layers (HKL) in the boundary regions of the GD [10]. In the case of ONO with HKL at the late stage of baking (in simulation procedure after ~ 10000 steps) we observed a sharp increase of the fraction of spreading electrons that results in a large increase of V_t shift. At the interface between the HKL and the nitride a large density of traps exists. These local centres trap electrons that lead to formation of additional charge at the interface. The newly injected carriers scatter on these charges and begin to move in a lateral direction. Later the number of spreading electrons can decrease due to the effect of competition between the formation of located charges in PP and those in the interface traps of the nitride. In Fig. 5 we demonstrate the behaviour of retention parameters of the memory cell with HKL that hinders the penetration of 90% of the IC outward from the ONO stack. One can see that at the first stages the existence of HKL improved retention characteristics of the memory cell. At some stage of baking the non-monotonous behaviour of the retention characteristics of the ONO stack with HKL is seen. Analysis of simulation results leads to the conclusion that the properties of the traps in the bulk and in the interface region are really responsible for this effect. Consequently in the process of memory device exploitation the effect of the non-monotonous kinetics of the retention characteristics of ONO memories with HKL can be expected. This effect depends also on the thickness of ONO (Fig. 6).

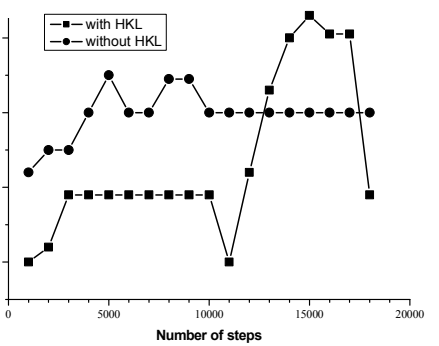


Figure 5: Two kinetics of IC in ONO without HKL and with HKL.

The vertical axis in Fig. 6 presents the number of steps preceding the jump on the kinetic curve shown in Fig. 5 for the case of ONO with HKL. The number of steps that corresponds to an increase of charge loss from the IR for the given depth of the ONO stack is larger for the HKL with smaller "k". This means that two limits of scaling down a memory cell exist. The upper one is caused by intensive scattering of injected carriers on trapping centers located at the interface between HKL and the nitride. The lower one is linked to charge leakage caused by tunneling current. The problem that should be solved is how to decrease the interval between these two limits.

The vertical axis in Fig. 6 presents the number of steps preceding the jump on the kinetic curve shown in Fig. 5 for the case of ONO with HKL. The number of steps that corresponds to an increase of charge loss from the IR for the given depth of ONO stack is larger for the HKL with smaller "k". This means that two limits of scaling down a memory cell exist. The upper one is caused by intensive scattering of injected carriers on trapping centers located at the interface between HKL and the nitride. The lower one is linked to charge leakage caused by the tunneling current. The problem that should be solved is how to decrease the interval between these two limits.

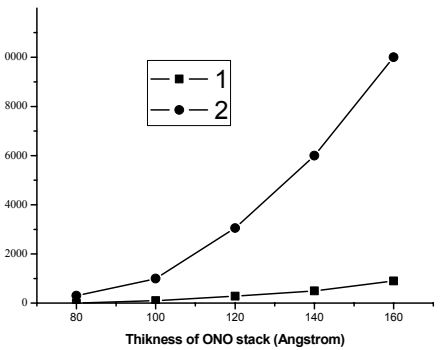


Figure 6: Influence of the quality of HKL on the possibility of scaling down of ONO stack. Graphs 1 and 2 correspond to HKL that hinders the ejection of 90% and 50% of IC from the ONO stack.

4 Conclusions

Novel software was used for simulation of physical processes responsible for V_t shift in a memory device. The proposed model of the GD accounts for both classical and quantum properties of the system. It was found that Coulomb repulsion determines the redistribution of the IC in the GD. As a result the nature of the V_t shift in cycled *micro*FLASH® memory transistors is explained and the ways for improvement of their parameters are indicated.

References

[1] Eitan B., Pavan P., Bloom I., Aloni E., Frommer A. & Finzi D., NROM: A novel localized trapping, 2-bit nonvolatile memory cell IEEE Elec. Dev. *Lett*., **21**, pp. 543-545, 2000.

[2] Roizin Y., Aloni E., Gutman M., Kairys V. & Zisman P., Retention characteristics of *micro*FLASH® Memory, *IEEE 2001 NVSMW*, Monterey, CA, 2001.

[3] Gritsenko V. A., Svitasheva S. N., Petrenko I. P., Wong H., Xu J. B. & Wilson I. H., Study of Excess Silicon at Si3N4/Thermal SiO_2 Interface Using EELS and Ellipsometric Measurements *J. Electrochem. Soc*. **146(2)**, pp. 780-785, 1999.

[4] Fuks D., Kiv A., Maximova T., Bibi R., Roizin Ya. & Gutman M., Computer model of the trapping media in microFLASH memory cells, *Journal of Computer-Aided Materials Design*, **9**, pp. 21-32, 2002.

[5] Yen C., Tsai W. J. & Lu T. C., Novel operation Schemes to Improve Device Reliability in a Localized Trapping Storage SONOS-type FLASH Memory, *Electron Devices Meeting, IEDM '03 Technical Digest. IEEE International*, pp. 7.5.1-7.5.4, 2003.

[6] Roizin Y., Yankelevich A. & Netzer Y., Novel technique for data retention and L_{eff} measurements in two bit microFLASH memory cells, *AIP Proceedings*, USA, pp. 181-185, 2001.

[7] Wang T., Tsai W. J., Gu S. H., Chan C. T., Yeh C. C., Zous N. K., Lu T. C., Pan S. & Lu C. Y., Reliability models of Data Retention and read-Disturb in 2-bit Nitride Storage FLASH Memory Cells, *Electron Devices Meeting, 2003, IEDM '03 Technical Digest. IEEE International*, pp.7.4.1-7.4.4, 2003.

[8] Fuks D., Kiv A., Roizin Ya., Gutman M., Bibi R. & Maximova T., The nature of HT Vt shift in NROM memory transistors, IEEE *Transactions on Electron Device*, **53**, pp. 304-313, 2006.

[9] Fuks D., Kiv A., Roizin Ya., & Gutman M., Computer simulation and experimental study of retention of SONOS device, *Computational Electronics*, **5**, pp. 49-52, 2006.

[10] Avichail-Bibi R., Kiv A., Maximova T., Roizin Y. & Fuks D., Behavior of injected electrons in high-k dielectric layers, *Materials Science in Semiconductor Processing*, **9**, pp. 985-988, 2006.

[11] MEDICI. User Guide, Version 2003.12. Synopsys, December 2003.

[12] Janai M., Data Retention, Endurance and Acceleration Factors of NROM Devices, *IEEE International Reliability Physics Seminar*, USA, pp. 502-503, 2003.

[13] Zisman P., Roizin Ya. & Gutman M., V_t drift of cycled two bit per cell microFLASH cells, *SSDM Proceedings*, Tokyo, Japan, pp. 228-230, 2003.

[14] Rudnikov T., Ostrovsky N., Fuks D., Kiv A., Bibi R., Roizin Ya. & Gutman M., Behavior of injected charges in dielectric layers of memory devices. *Proceedings of the 2nd International Conference on Information Technologies and Management*, Riga, Latvia, pp.69-74, 2004.

[15] Avichail-Bibi R., Fuks D., Kiv A., Roizin Y. Maximova T., Gouternakht A. & Shtermer V., Software for simulation of retention loss in memory transistor, *Computer Modeling & New Technologies,* **10**(2), pp.30-39, 2006.
[16] Mott N. F. & Devis E. A., *Electron processes in non-crystal materials*, Clarendon Press, Oxford, 1979.