Monte Carlo Simulation of Charge Carrier Injection in Twin Flash™ Memory Devices during Program and Erase

R. Hagenbeck[1], S. Decker[2], P. Haibach[1]
[1]Qimonda, Munich, Germany
[2]Infineon Technologies, Munich, Germany
rainer.hagenbeck@qimonda.com

T. Mikolajick[3], G. Tempel[4], M. Isler[3]
[3]Qimonda, Dresden, Germany
[4]Infineon Technologies, Dresden, Germany

C. Jungemann
Bundeswehr University Munich, Germany

B. Meinerzhagen
Technical University Braunschweig

***Abstract*—An iterative and time-dependent simulation method based on a full band Monte Carlo algorithm is presented to describe the injection behavior of hot electrons and holes during program and erase of Twin Flash™ memory cells. Secondaries during programming and the feedback of already injected and trapped charge carriers in the ONO nitride on subsequent injection processes are taken into account. By this method it is possible to obtain valuable information on the time-dependent evolution and the local distribution of injection currents and trapped charges in the ONO nitride of the Twin Flash™ cell.**

Keywords: NROM, Twin Flash memories, hot carrier injection, secondaries, program, erase, Monte Carlo simulation.

I. INTRODUCTION

Twin Flash™ memory cells are programmed and erased by the localized injection of hot charge carriers into the nitride layer of an ONO stack (Fig. 1) [1]. The localization of the trapped charges in the insulating nitride layer close to the drain or source junction of the transistor-like memory device allows the storage of two separated bits in each cell. For an efficient cell optimization the information is needed where exactly the injection of charged species takes place and how local distributions of injection current and trapped nitride charge evolve during the time frame of single program and erase steps and also after several programming and erase cycles. In order to obtain reliable simulation results it is important to take into account the feedback of already injected and trapped carriers on the distribution of the local field and the injection current. The details of the applied iterative simulation method are described in [3]. The time- and cycling-dependent evolution of the trapped charge carrier distribution in the insulating nitride layer of the ONO has significant impact on the reliability behavior of the memory cell which comprises endurance, retention after cycling, and the separation of the two bits stored in every cell. Therefore, a Monte Carlo based simulation setup was developed to study the details of the time-dependent hot electron (program) and hot hole (erase) injection in Twin Flash™ memory cells.

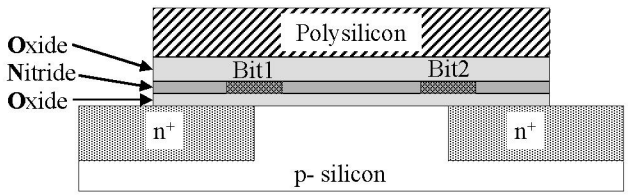

Figure 1. Simulated 2d structure of Twin Flash™ memory cell.

Under programming conditions, hot channel electrons generated by the high voltage between source and drain are injected into the ONO nitride close to the drain junction of the cell device. As a consequence, the reverse read threshold voltage in the saturation regime increases compared to the virgin state. The modification of the forward read threshold voltage has to be as low as possible to guarantee for a good 2 bit separation. For erase, the generation mechanism of hot holes is band-to-band-tunneling close to the drain junction. Those hot holes which overcome the potential barrier of 4.1eV at the silicon/silicon oxide interface are also injected into the nitride layer. There, they have to compensate the negative charges originating from the programming process in order to reduce the reverse threshold voltage again. It is obvious that the knowledge of the local injection current distribution during program and erase is very important for an optimized memory cell design. The electron and hole injection currents are calculated by applying a full band Monte Carlo (MC) method [2] in a post processing mode instead of the non-local lucky electron model mentioned in [3]. The electric potential and field distribution originates from a classical hydrodynamic simulation carried out by the device simulator GALENE [2][4].

All charge carriers which overcome the interface between silicon substrate and bottom oxide are assumed to be injected into the traps of the ONO nitride. The details of the oxide transport as well as the influence of oxide damage due to the injection processes are not taken into account in our simulations. The injected charge carriers are assumed to be distributed homogeneously along the vertical layer thickness of the nitride. A lateral charge carrier transport within the nitride is also neglected. These missing model aspects will be implemented in the future. Nevertheless, the presented simulation method gives valuable insight into the details of injection phenomena during program and erase of Twin Flash™ memory cells leading to a better understanding and optimization of the memory device performance.

This work was financially supported by the Federal Ministry of Education and Research of the Federal Republic of Germany (Project No. 01M3160). The authors are responsible for the content of the paper.

II. Simulation Model of Hot Carrier Injection by Full Band Monte Carlo Method

A. Electron Injection, Secondaries

Electrons are simulated by the full band Monte Carlo model described in Ref. [2], which includes scattering by phonons and impurities and impact ionization. Electrons, which impinge on the silicon/oxide interface, can be injected into the oxide, if the total energy and parallel momentum can be conserved, where a simple parabolic band centered at the Γ-point with a mass of 0.5 times the free electron mass is assumed for the oxide. The height of the barrier is given by the maximum of the electrostatic potential minus the image potential (Schottky barrier lowering) plus the conduction band shift between silicon and oxide. The maximum is searched for in the direction perpendicular to the channel over the whole thickness of the oxide layer. A conduction band shift of 3.1eV is used and the parameter of the Schottky barrier lowering was determined by matching oxide injection experiments [8]. Electrons with an energy below the barrier can tunnel into the nitride with a probability given by the WKB formula evaluated for the potential profile perpendicular to the channel at the injection point. The poor statistics of oxide injection are enhanced by the multiple refresh method [2].

Impact ionization of hot channel electrons near the drain of an NMOSFET generates electron/hole pairs. The holes are accelerated towards the substrate and can generate secondary electron/hole pairs if they gain sufficient energy [6]. These generated electrons are the so-called secondaries, which have a lower potential energy than electrons injected from the sources due to the built-in voltage of the drain/bulk junction. This leads to a higher injection probability of these electrons and a spatial injection distribution, which differs from the one of the channel hot electrons [7].

B. Hole Injection

The hole model is similar to the electron model and includes the same scattering mechanisms. The parameters of the hot hole injection model are determined by matching experimental results of Ref. [9] and the hole mass in the oxide is assumed to be equal to the one of the electrons.

III. Simulation Results

A. Injection behavior of virgin cell

First MC simulations of hot electron injection during program and hot hole injection during erase of virgin Twin Flash memory cells were done by Ingrosso et al. [5]. They used a template device with an effective channel length of 0.35µm and an idealized dopant distribution. For our MC simulations, we used realistic dopant distribution and device topology which was calculated by the process simulation tool TSUPREM4. Fig. 2 and 3 show the simulated local distribution of electron injection current during programming of a virgin Twin Flash™ cell for different gate and drain voltages. The maximum of the electron injection current is located at the position of the drain junction at the oxide interface. Electrons are injected into the nitride out of the channel region as well as out of the n^+ drain region. The gate voltage has only a slight

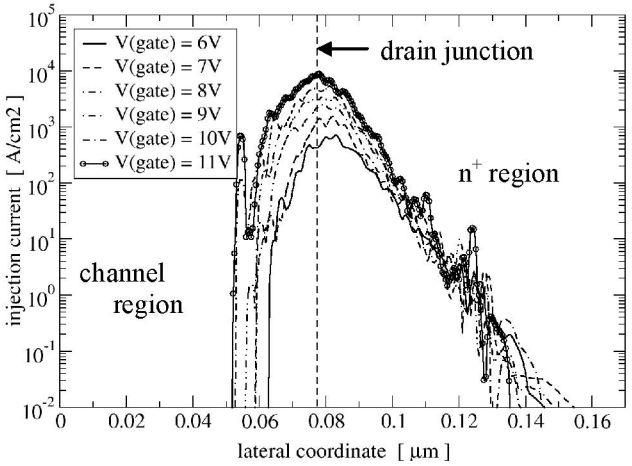

Figure 2. Local distribution of electron injection current at programming conditions for different gate voltages, drain voltage = 4V, gate length = 230nm.

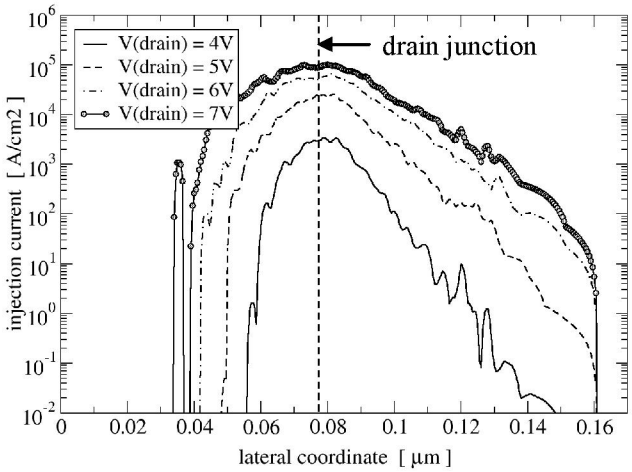

Figure 3. Local distribution of electron injection current at programming conditions for different drain voltages, gate voltage = 9V , gate length = 230nm.

impact on the position of the injection current maximum and the extension of the injection region into the channel (Fig. 2). A higher drain voltage enlarges the absolute value of the injection current as well as the width of the injection region (Fig. 3).

The shift of the reverse read threshold voltage is mainly determined by the portion of hot carrier injection in the channel region. The width of the injection current distribution in the channel region varies from 15 to 40nm depending on the programming voltage conditions. Simulations of smaller device geometries with reduced effective channel length indicate that in this case the width of the injection region is also reduced.

Fig. 4 and 5 show the gate and drain voltage dependence of the hole injection current at overerase of a virgin cell. During erase, hot holes are exclusively injected in the channel region. No hole injection happens at the n^+ drain region,. The lateral width of the hole injection region is nearly independent of the gate voltage but strongly influenced by the drain voltage. By comparing the hole injection current distributions of gate/drain

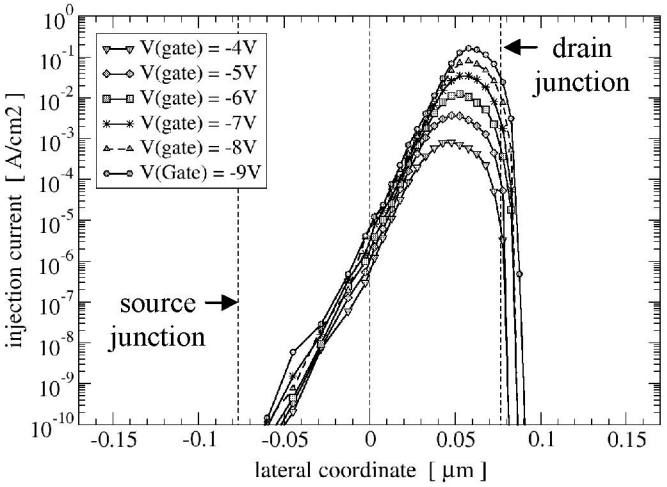

Figure 4. Local distribution of hole injection current at erase conditions for different gate voltages, drain voltage = 5V, gate length = 100nm.

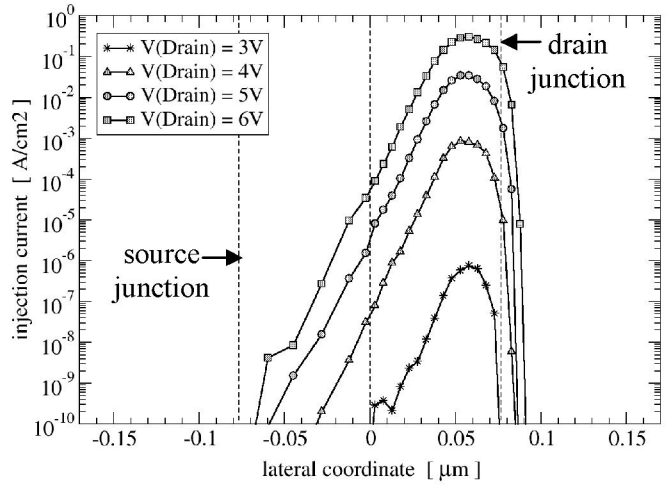

Figure 5. Local distribution of hole injection current at erase conditions for different drain voltages, gate voltage = -7V, gate length = 100nm.

bias combinations of -9V/+5V and -7V/+6V it becomes clear that a more positive gate bias during erase leads to a broader injection of holes in the nitride. This would facilitate the compensation of secondary electrons injected over the channel (see below). This conclusion is very important for improving the reliability of Twin Flash™ cells with channel lengths below 100nm.

Besides the injection of primary hot electrons during program, secondarily generated electrons (due to impact ionization by hot holes in the bulk region caused by channel hot electrons) are also taken into account [6][7]. The secondaries become important for negative bulk voltages during program. Fig. 6 shows the increasing amount of injected secondaries with increasing negative bulk bias during program of a virgin cell. For zero bulk bias, secondaries do not contribute significantly to the overall injection current whereas for strongly negative bulk bias they are injected over the whole channel region degrading the separation of the two bits stored in the cell.

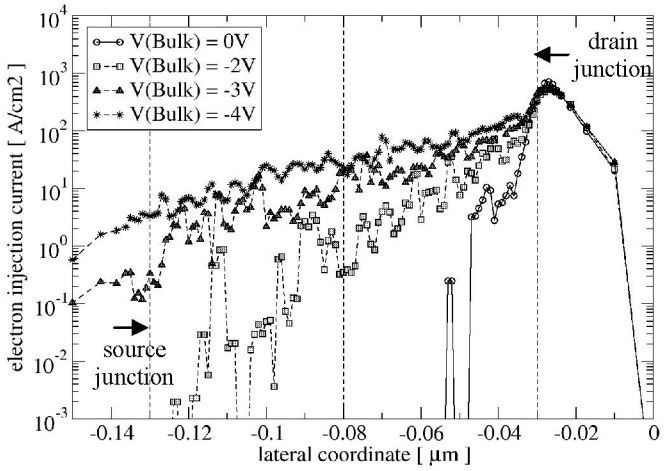

Figure 6. MC simulation of a virgin cell, local distribution of electron injection current including secondaries under programming conditions, different bulk voltages, V(gate) = 5V, V(drain) = 3.5V, gate length = 100nm.

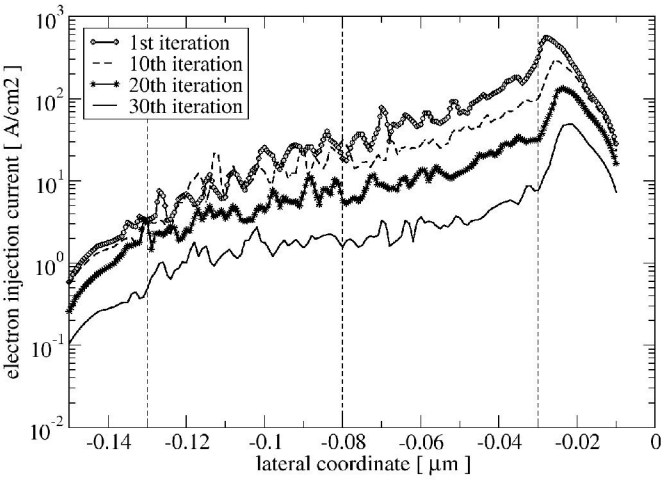

Figure 7. Spatial distribution of electron injection current at different iteration/time steps of programming including secondaries. Applied programming conditions: gate voltage = 5V, drain voltage = 3.5V, bulk voltage = -4V. The dashed vertical lines indicate the position of source and drain junction and the centre of the channel.

B. Time dependence of carrier injection

Using the initial programming state of a virgin cell as a starting point, an iterative algorithm can be applied to calculate the time dependence and the local distribution of the injection current taking into account the feedback of already injected and trapped nitride charges on subsequent injection processes. The details of this method are described in [3]. As an example, Fig. 7 illustrates the time-dependent evolution of the spatial distribution of the electron injection current during programming at different iteration steps of the simulation procedure. In this example, a very high bulk bias of -4V is applied to emphasize that the feedback effect of already trapped nitride charges is also reducing the part of the injection current which is caused by secondaries. Each iteration step corresponds to a distinct time and is defined by a constant increase of the injected nitride charge. The time interval corresponding to a special iteration is

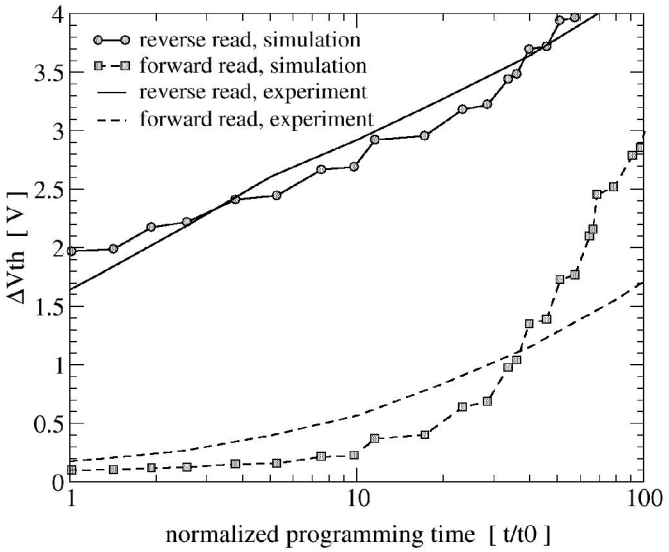

Figure 8. Time dependence of threshold voltage shift ΔV_{th} during programming for reverse and forward read. Comparison between experiment and iterative Monte Carlo simulation, gate length = 230nm.

determined by the ratio of the overall injection current at this iteration step and the predefined amount of injected charge per iteration step [3]. With increasing programming time the electron injection current is decreasing due to the electrostatic feedback of already injected and trapped nitride charges on the field distribution in the silicon and the bottom oxide of the ONO layer (Fig. 7). The injection of hot electrons into the ONO nitride layer becomes more difficult. Fig. 8 shows a comparison between the iterative MC simulation and experimental data under programming conditions for a Twin Flash™ cell with a gate length L_g of 230nm. The shift of the threshold voltages ΔV_{TH} in forward and reverse read direction is plotted as a function of program-time of bit 2. The calibration of the simulation to experiment is done by introducing a limit on the trap density in the nitride layer of the ONO stack. The carriers which are injected into a nitride region where all nitride traps are already filled have to be transferred into free traps in lateral direction. This lateral redistribution leads to the strong increase of both the threshold voltage in forward and reverse direction due to the pronounced increase of trapped electrons over the channel region.

IV. Conclusions

In conclusion, we have shown MC simulation results which give detailed insight into the local distribution of the hot carrier injection for program and erase of Twin Flash™ memory cells. For the first time, the local distribution of injection of hot holes during erase was simulated by an MC method. In addition, a model for the generation and subsequent injection of secondarily generated electrons during program was applied. Since the injected charges themselves influence the characteristics of subsequent injection processes, it is necessary to simulate iteratively and self-consistently the evolution of injection current and trapped nitride charge during programming. In this way, the presented simulation method helps to find optimal voltage conditions for program and erase for different cell geometries and fabrication processes

References

[1] B. Eitan, P. Pavan, I. Bloom, E. Aloni, A. Frommer, and D. Finzi., „NROM: A novel localized trapping, 2-bit non-volatile memory cell," IEEE Electron Device Letters, vol. 21, no. 11, 2000, pp. 543-545.

[2] C. Jungemann and B. Meinerzhagen, Hierarchical Device Simulation, Springer Verlag Wien/New York, 2003.

[3] R. Hagenbeck, S.Decker, F. Lau, P. Haibach, J.-M. Schley, M. Isler, T. Mikolajick, and G. Tempel, „Modeling and simulation of electron injection during programming in Twin Flash™ devices based on energy transport and the non-local lucky electron concept," J. Comp. Electronics, vol. 3, 2004, pp. 239-242.

[4] B. Meinerzhagen and W. L. Engl, "The influence of the thermal equilibrium approximation on the accuracy of classical two-dimensional numerical modeling of silicon submicrometer MOS transistors", IEEE Trans. Electron Devices, vol. 35, no. 5, 1988, pp. 689-697.

[5] G. Ingrosso, L.Selmi, and E. Sangiorgi, "Monte Carlo simulation of program and erase charge distributions in NROM™ devices", Proceeding of ESSDERC 2002, pp. 187-190.

[6] J. D. Bude, "Monte Carlo simulation of impact ionization feedback in sub-micron MOSFET technologies," Ext. Abst. of the 1995 Int. Conf. on Solid State Devices and Materials, 1995, pp. 228-230.

[7] C. Jungemann, S. Yamaguchi, and H. Goto, "Investigation of the influence of impact ionization feedback on the spatial distribution of hot carriers in an N-MOSFET," Proceeding of ESSDERC 1997, p. 336-340.

[8] T. H. Ning, C. M. Osburn, and H. N. Yu, "Emission probability of hot electrons from silicon into silicon dioxide," J. Appl. Phys., vol. 48, 1977, pp. 286-293.

[9] L. Selmi, E. Sangiorgi, R. Bez and B. Ricco, "Measurement of the hot hole injection probability from Si into SiO_2 in p-MOSFET's," IEDM Tech. Dig., 1993, pp. 333-336.