

J-2-4

A Study on Scaling Limit of Programming Characteristics of NOR-Flash Memories using Device Simulation

Takamitsu Ishihara¹, Takao Marukame¹, Kazuya Matsuzawa¹, and Takahiro Nakauchi²

¹Advanced LSI Technology Laboratories, Corporate Research & Development Center, Toshiba Corporation
8, Shinsugita-cho, Yokohama, Kanagawa 235-8522, Japan
Phone: +81-45-776-5937 E-mail: takamitsu.ishihara@toshiba.co.jp

²Memory Div., Mobile Memory Device Engineering Dept., Semiconductor Company, Toshiba Corporation,
2-5-1 Kasama, Sakae-ku, Yokohama, Kanagawa 247-8585, Japan

1. Introduction

Owing to the increasing demand for high-density, low-power and low-cost electrically re-programmable nonvolatile memories, aggressive device size scaling of NOR-Flash memories is desired. Channel hot electron (CHE) injection is the conventional method to program the floating gate (FG). However, as a consequence of the aggressive device scaling, degradation of programming characteristics using the CHE injection has become a serious problem for scaled NOR-Flash memories [1]. The purpose of the present paper is to clarify the degradation mechanism and the scaling limit of programming characteristics using the CHE injection of scaled NOR-Flash memories based on device simulation.

2. Simulation Method

We have used the energy transport model (ETM) to evaluate the programming characteristics of NOR-Flash memories because ETM can properly take into account the hot-carrier-related phenomena (Figure 1). The distribution function of hot electrons (HEDF) is also an important quantity in evaluating the programming characteristics by CHE injection. Though the actual functional form of HEDF is complicated [2], we have found that the following simple form can describe well the hot-carrier related phenomena.

$$f(\varepsilon) = \exp\left(-a\left(\frac{\varepsilon}{kT}\right)^b\right) \quad (1)$$

We have determined the parameters $a=0.5$ and $b=1.5$ by reproducing the experimental substrate current, as shown in Figure 2. The gate current injected into the FG has been evaluated by solving the ETM and the current continuity equation in the gate insulator, by taking the electron density and the average electron energy injected into the gate insulator as the fixed boundary condition, as shown in Fig. 1. Though the FG was treated as metal, a voltage drop in the FG was taken into account by depletion approximation [3].

3. Calculated Results

The Device structure used in the simulation is shown in Figure 3. To realize the quantitative estimation, the experimental current-voltage and the programming characteristics need to be reproduced accurately by device simulation. As shown in Figure 3, the experimental threshold voltages for various gate lengths have been reproduced well by taking into account the accurately evaluated substrate impurity profile by in-house process simulator, TOPAZV2.

The input pulse-form of the control gate (CG) and drain voltages used in the simulation of the programming characteristics is shown in Fig. 3. The drain voltage and the CG

voltage dependence of the programming characteristics limited by CHE injection has been represented well by our simulation, as shown in Figures 5 and 6. Thus, the accuracy of our simulation has been justified.

Based on the accurate device simulation thus constructed, we have studied the scaling limit of CHE injection in programming operation. As shown in Figure 7, the increase of a voltage drop in the FG caused by the CHE injection with the decrease in the gate length has been observed for the gate length greater than 130nm. This is due to the increase of the drain current (Figure 8), because drain voltage is taken to be 3.5V for all gate lengths and the electric field along the channel becomes strong as the gate length decreases. On the other hand, the increase of a voltage drop in the FG caused by the CHE injection was not observed for the gate length of less than 130nm (Fig. 7). This behavior cannot be understood in terms of the increase of the drain current with the decrease of the gate length (Fig. 8). To gain the insight into this discrepancy, we have examined the behavior of electron velocity near the drain region, as shown in Figure 9. From Fig. 9, we have found a remarkable electron velocity overshoot near the drain region. This electron velocity overshoot near the drain region becomes enhanced as the gate length decreases. The enhanced electron velocity overshoot means the increase of ballistic hot electrons injected into the drain region without injection into the FG. Thus, we have found that the electron velocity overshoot effect severely limits the efficiency of the CHE injection in programming operation of scaled NOR-Flash memories.

4. Conclusions

We have constructed a device simulation technique that enables the accurate evaluation of the programming characteristics using the CHE injection by solving the ETM and the current continuity equation in the gate insulator. It has been found that the programming characteristics due to the CHE injection significantly degrade because of the enhanced velocity overshoot effect for scaled NOR-Flash memories with gate length less than 130nm.

Acknowledgements

We would like to express sincere thanks to Mr. Hiroyuki Sasaki for his support of our study.

References

- [1] J. D. Bude, M. R. Pinto, and R. K. Smith, ED47 (2000) 1873.
- [2] K. Sonoda et al., J. Appl. Phys. **80** (1996) 5444.
- [3] K-F. Schuegraf and C. Hu, ED41 (1994) 761.

Energy transport model

Current continuity equation

$$\frac{\partial n}{\partial t} + \nabla \cdot n \mathbf{v} = U, \quad n \mathbf{v} = \mu \left(n \nabla \psi - \frac{kT}{q} \nabla n - \frac{kn}{q} \nabla T \right)$$

Energy conservation equation

$$\frac{\partial mnw}{\partial t} + \nabla \cdot m \mathbf{S} - qmn \mathbf{v} \cdot \nabla \psi = mwU - n \frac{mw - m_0 w_0}{\tau_w}$$

v: drift velocity

Poisson equation

$$\nabla[\epsilon \nabla \psi] = q(n - p + N_a - N_d)$$

n: electron density

ψ : potential

T: electron temperature

m: electron effective mass

w: electron energy

q: elementary charge

U: generation-recombination rate

$$m \mathbf{S} = \frac{5}{2} \mu \left(mnkT \nabla \psi - \frac{k^2}{q} \nabla mnT^2 \right)$$

$$w = \frac{3}{2} kT + \frac{1}{2} m(w)v^2$$

ϵ : dielectric constant

μ : electron mobility

$$E_{\text{flux}} = \frac{1}{2} \mu \left(mnkT \nabla \psi - \frac{k^2}{q} \nabla mnT^2 \right)$$

$$w = \frac{3}{2} kT + \frac{1}{2} m(w)v^2$$

Fig. 1 Explanation of energy transport model and gate injection current model used in this study.

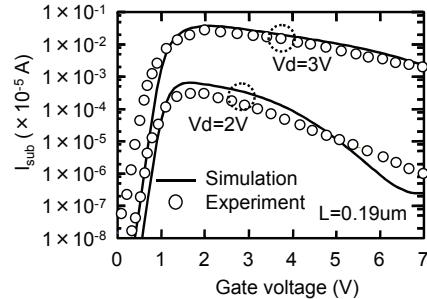


Fig. 2 Comparison of the experimental and the simulated I_{sub} (substrate current) as a parameter of drain voltage.

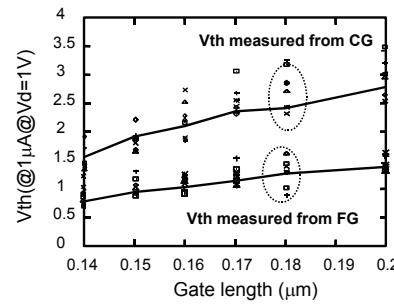


Fig. 4 Comparison of the experimental (symbols) and the simulated V_{th} (solid lines).

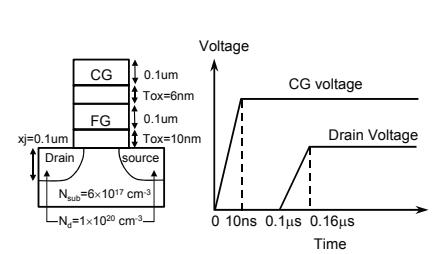


Fig. 3 Device structure and the wave form of the input control gate and drain voltage.

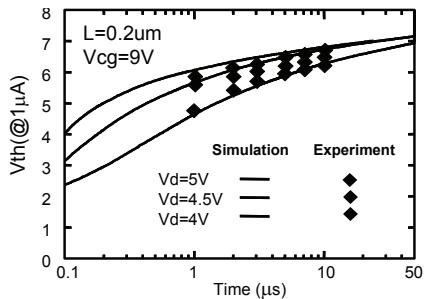


Fig. 5 Comparison of the experimental and the simulated V_{th} after program operation as a parameter of drain voltage.

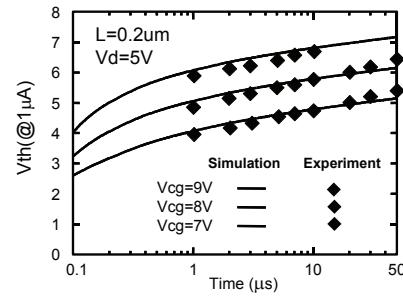


Fig. 6 Comparison of the experimental and the simulated V_{th} after program operation as a parameter of control gate voltage.

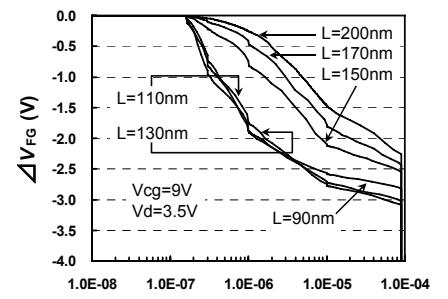


Fig. 7 Voltage drops in floating gate caused by hot electron injection as a parameter of gate length.

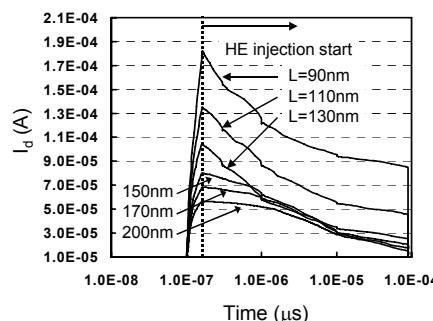


Fig. 8 Drain current (I_d) during program operation as a parameter of gate length.

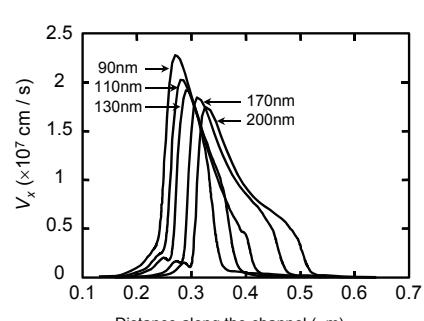


Fig. 9 Velocity over-shoot at the surface region in substrate as a parameter of gate length.