Pergamon

0038-1101(94)00087-5

Solid-State Electronics Vol. 38, No. 2, pp. 461–464, 1995

0038-1101/95 $9.50 + 0.00

TRANSIENT SIMULATION OF EPROM WRITING CHARACTERISTICS, WITH A NOVEL HOT ELECTRON INJECTION MODEL

CHIMOON HUANG and TAHUI WANG

Department of Electronics Engineering, Institute of Electronics, National Chiao-Tung University, Hsin-Chu, Taiwan, Republic of China

(Received 7 October 1993; in revised form 23 April 1994)

Abstract—A two-dimensional transient simulation of EPROM writing characteristics is presented. A Monte Carlo-based hot electron injection model which accounts for Fowler–Nordheim tunneling and thermionic emission has been included in the simulation. The simulated EPROM writing transient characteristics is compared favorably with experimental results for channel lengths down to 0.5 μm. The importance of the two injection mechanisms, thermionic emission and quantum tunneling, is evaluated.

1. INTRODUCTION

Although channel hot-electron-injection (CHEI) into a gate can produce serious reliability problems in short channel n-MOSFETs, this approach has been widely utilized in erasable programmable read-only memory (EPROM) cells due to its compatibility with standard MOS processes and high reliability for long channel devices. With an increase of the memory integration level[1], improvements of the programming time in today's design of EPROM cell structures become of particular interest. To facilitate the device design, an accurate and predictive simulator which is able to model oxide tunneling current and thermionic emission current is needed.

An accurate simulation of hot electron injection in EPROM devices is complicated since the calculation of the hot electron distribution requires detailed knowledge about electron transport in a complex Si band-structure. Previous EPROM modeling methods were mostly based on a lucky-electron concept with various fitting parameters[2]. Since the formulation of the lucky-electron model does not account for the important non-local heating effect, Goldsman and Frey were the first to use a combined hydrodynamic model and a non-Maxwellian electron energy distribution to simulate hot electron injection in MOSFETs[3]. Recently, Cassi and Ricco have derived a closed-form hot electron distribution from the first-principle Boltzmann transport equation based on assumptions of a single conduction band, a single scattering mechanism (optical phonon scattering), and relatively slow variation of electric field in space[4]. Subsequently, Fiegna *et al.* used this model and proposed a simple and efficient numerical technique to analyze the EPROM cell writing characteristics[5]. Unfortunately, these approaches either need to use some unphysical parameters which may severly limit the validity and the application of the models, or have to make some assumptions which are hardly justified in the deep submicron domain.

More recently, we have developed a sophisticated method to determine the high energy tail of the channel hot electron distribution along the Si/SiO_2 interface from an ensemble Monte Carlo (EMC) simulation including a complete pseudopotential band-structure[6]. The hot electron gate current MOSFETs therefore can be calculated without using fitting parameters. In this work, we extend the Monte Carlo-based hot electron model to the simulation of EPROM writing transient characteristics.

2. PHYSICAL MODEL AND SIMULATION METHOD

2.1. Transient simulation

The simulation procedure of EPROM writing characteristics is described in Fig. 1. A two-dimensional device simulator, Silvaco PISCES IIB, is used to calculate the device electrical characteristics within a hydrodynamic energy transport frame. The boundary conditions in PISCES IIB are appropriately modified to incorporate a floating gate. The charging current to a floating gate via the CHEI is then evaluated in each time step using the rigorous Monte Carlo method as a post-processor. The evaluation of the hot electron injection current in a window Monte Carlo simulation will be described in Section 2.2. The two-dimensional electric field profile in the Monte Carlo simulation is generated from PISCES IIB.

The programming operation, starting at charging time $t_0 = 0$ with an initial floating gate charge $Q_0 = 0$, is divided into time steps of variable length. In each time step, the accumulation of stored charge in a floating gate is ΔQ_{fg}. If the charging current is large,

the time step is chosen smaller to maintain a constant ΔQ_{fg}. In this way, we can control the numerical error associated with time discretization. The charging current I_g is assumed to be constant in each time step. Therefore, the relationship between writing time t and floating gate charge Q_{fg} is readily expressed as:

$$t = t_0 + \sum_i |\Delta Q_{fg}|/(I_g)_i. \quad (1)$$

It should be mentioned that the thickness of a floating gate is generally larger than 0.1 μm. The control gate leakage current caused by hot electron ballistic motion through the floating gate is extremely small. The assumption of no leak of injected electrons in eqn (1) during the programming is reasonable.

The threshold voltage shift $(\Delta V_{th})_i$ caused by the CHEI in the ith time interval in a cell transistor is simply calculated as[7]:

$$(\Delta V_{th})_i = -(Q_{fg})_i t_{ox2}/\epsilon_{ox}, \quad (2)$$

where $(Q_{fg})_i = Q_0 + \Sigma \Delta Q_{fg}$ and t_{ox2} represents the thickness of the oxide layer between the control gate and the floating gate. The above iteration is continued until the charging current is less than a certain value.

2.2. Hot electron injection

The CHEI current in a MOS structure is expressed by[6]:

$$I_g = I_{ds} \int_{channel} \int_0^\infty P_c(x) f(x, E) D(E) P_t(x, E) dE \, dx, \quad (3)$$

where x is along the channel direction, $D(E)$ is the electron density of states of a realistic Si pseudopotential band-structure, $P_c(x)$ denotes the probability that electrons impinge on the Si/SiO_2 interface when they travel in the channel. $P_t(x, E)$ is the electron transmission probability across the oxide which involves the contribution from both quantum tunneling and thermionic emission mechanisms. The tunneling probability is evaluated from the WKB method with inclusion of the Schottky lowering effect. Electron transport with energies above the SiO_2 barrier is considered as thermionic emission. The current loss due to phonon scattering in the oxide layer is also taken into account. $f(x, E)$ represents the electron distribution function, which is obtained from an electron energy distribution divided by $D(E)$. Since the CHEI is caused by electrons with energies around the Si/SiO_2 interface barrier height (3.0 eV), $f(x, E)$ cannot be resolved completely from a Monte Carlo simulation in the entire energy region of interest due to limited computational resources. In this work, we calculate the electron energy distribution and the $f(x, E)$ directly from a Monte Carlo method only at low and intermediate electron energies. In the portion of high energy, an extrapolation is adopted with a particular hot electron temperature extraction technique:

$$\langle E(x) \rangle = \frac{\int_{E_1}^\infty E \exp(-E/kT_e) D(E) \, dE}{\int_{E_1}^\infty \exp(-E/kT_e) D(E) dE}, \quad (4)$$

where E_1 is the onset of the extrapolation and $\langle E(x) \rangle$ represents an energy average for those electrons with energies above E_1. With $\langle E(x) \rangle$ computed from the Monte Carlo simulation, one can easily extract hot electron temperature T_e from eqn (4). The choice of the integration lower limit E_1 is considered in two aspects: if E_1 is chosen too large, the statistical error when computing $\langle E(x) \rangle$ increases due to a reduced number of sample electrons above E_1 in the Monte Carlo simulation; on the other side, if E_1 is chosen too small, the extracted T_e cannot appropriately reflect the energy distribution in the energy range of interest. In the present simulation, E_1 is 2.0 eV. It should be pointed out that the validity of the above extraction technique is based on the simulation result that although $f(x, E)$ shows a markedly non-Maxwellian feature globally at a high field, it still exhibits an approximately exponential decay in an energy region much above the average electron energy.

3. RESULT AND DISCUSSION

Figure 2 shows the simulated and measured gate current as a function of a gate bias for a MOSFET

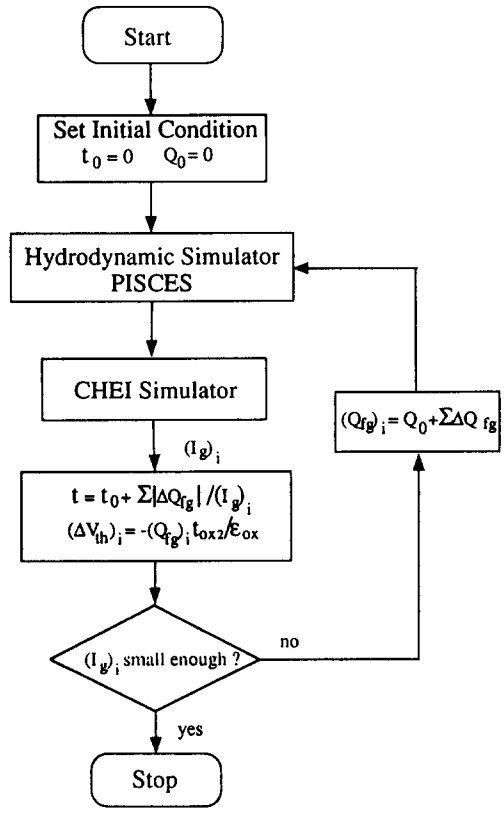

Fig. 1. Simulation flowchart of EPROM writing characteristics.

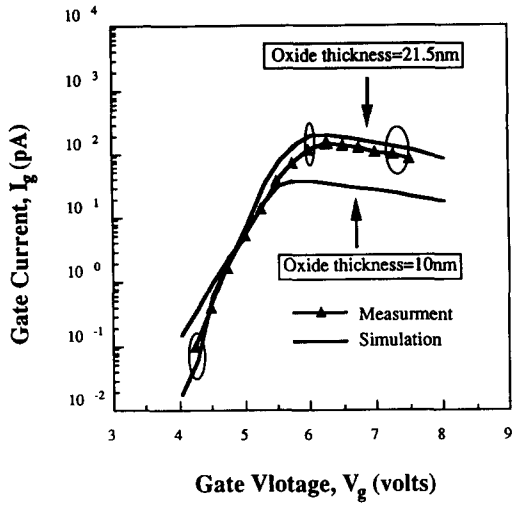

Fig. 2. Comparison of calculated and measured gate currents vs a gate bias.

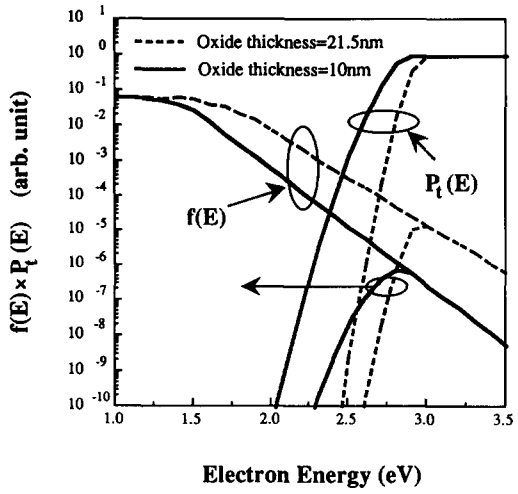

Fig. 3. Product of electron distribution function $f(E)$ and oxide transmission probability $P_t(E)$ as a function of electron energy. V_{ds} is 5.5 V and V_{gs} is 7 V.

with an effective channel length of 0.8 μm at a drain voltage of 5.5 V. The gate bias ranges from 4 to 8 V, a typical region for EPROM writing. The device specifications are in the following: n^+ source/drain surface doping = 4×10^{20} cm^{-3}; channel surface doping = 1.7×10^{17} cm^{-3}; junction depth = 0.25 μm; gate oxide thicknesa = 215 Å. Good agreement between the calculated gate current and the experimental result[5] in Fig. 2 verifies our method.

The calculated gate current in a 100 Å gate oxide device is also shown in Fig. 2 for a comparison. It should be pointed out that the 100 Å gate oxide device has a larger gate charging current only at low gate biases. It implies that the reduction of gate oxide thickness cannot effectively improve the EPROM programming speed. A similar measurement result was reported by Toyoshima *et al.*[8] for 50 and 100 Å devices. In order to explain the result, the product of two important terms in eqn (3), electron distribution function $f(x, E)$ and oxide transmission probability $P_t(x, E)$, at the position of maximum hot electron injection is plotted as a function of electron energy in Fig. 3. The gate bias is 7.0 V. Although the 100 Å device has a larger tunneling probability, a higher electron temperature [indicated by the slope of $f(x, E)$] due to a larger channel field in the 215 Å device results in a higher gate charging current. In addition, while the dominant injection mechanism in the 100 Å device is quantum tunneling, Fig. 3 shows that a large fraction of the product of $f(x, E)$ and $P_t(x, E)$ in the 215 Å device is above the Si/SiO_2 interface barrier height. This suggests that thermionic emission rather than quantum tunneling is a major injection mechanism in the 215 Å device.

The schematic representation of simulated EPROM structures is illustrated in Fig. 4. A Monte Carlo window is positioned to include all possible CHEI trajectories. In the inset of the figure is the sideview of the device. An additional 0.7 μm overlap region between the control gate and the floating gate is used to increase the coupling capacitance. In the transient simulation, smaller time steps are used in the initial charging period when I_g is high. A larger time step is then adopted while the injection current is low. The calculated writing transients for a 0.7 μm channel length at a drain bias V_{ds} of 5.75 and 6.5 V are shown in Fig. 5. The solid lines in the figure are the measured result[5]. Initially, the floating gate potential is much higher than V_{ds} and the threshold voltage changes rapidly due to a high injection current. Then, the writing current decreases abruptly as a result of electron accumulation in the floating gate. The threshold voltage shift tends to saturate as the floating gate potential is lower than V_{ds}. In Fig. 6, the

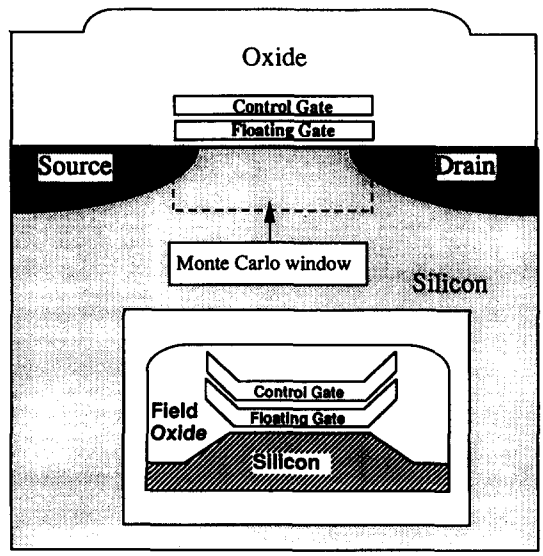

Fig. 4. Schematic representation of simulated EPROM structures and a Monte Carlo simulation window. In the inset of the figure is the sidewiew of the device.

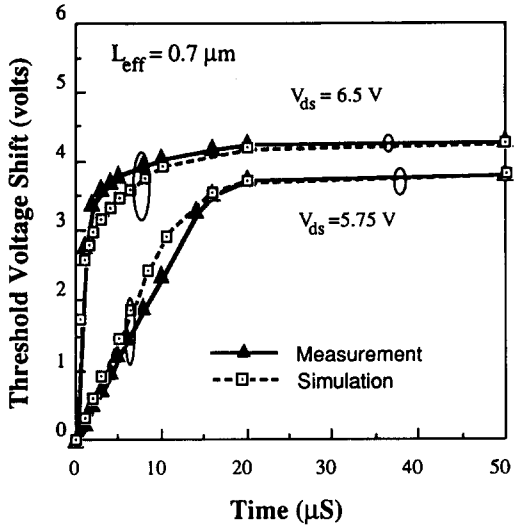

Fig. 5. Comparison of simulated and measured writing transient characteristics.

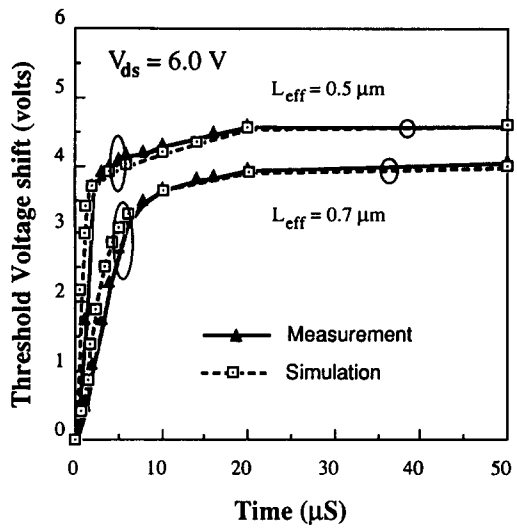

Fig. 6. Simulated and measured writing transient characteristics for 0.5 and 0.7 μm devices.

writing transient characteristics for 0.5 and 0.7 μm EPROM cells are drawn. Our simulation shows a good agreement with the experimental data[5].

4. CONCLUSION

In conclusion, we have developed a simulation technique to analyze the writing mechanisms in the EPROM devices. The effects of both thermionic emission and quantum tunneling have been evaluated. The accuracy of our model has been verified by good consistency between the simulated and measured results for channel lengths down to 0.5 μm.

Acknowledgement—Financial support from National Science Council, ROC, under contract No. NSC81-0404-E-009-106 is greatly acknowledged.

REFERENCES

1. N. Ajika, M. Ohi, H. Arima, T. Matsukawa and N. Tsubouchi, *IEDM Tech. Dig.*, pp. 115–118 (1990).
2. C. Hu, *IEDM Tech. Dig.*, pp. 22–25 (1979).
3. N. Goldsman and J. Frey, *Solid-St. Electron.* **31,** 1089 (1988).
4. D. Cassi and B. Ricco, *IEEE Trans. Electron Devices* **ED-37,** 1514 (1990).
5. C. Fiegna, F. Venturi, M. Melanotte, E. Sangiorgi and B. Ricco, *IEEE Trans. Electron Devices* **ED-38,** 603 (1991).
6. C. Huang, T. Wang, C. N. Chen, M. C. Chang and J. Fu, *IEEE Trans. Electron Devices* **ED-39,** 2562 (1992).
7. B. Eitan and D. Frohman-Bentchkowsky, *IEEE Trans. Electron Devices* **ED-28,** 328 (1981).
8. Y. Toyoshima, H. Iwai, F. Matsuoka, H. Hayashida, K. Maeguchi and K. Kanzaki, *IEEE Trans. Electron Devices* **ED-37,** 1496 (1990).