Concurrent Technology, Device, and Circuit Development for EEPROMs

U. Feldmann*, R. Kakoschke**, M. Miura-Mattausch***, and G. Schraud**
* Corporate Technology, ** Semiconductor Division
ZT PP 2 S, Siemens AG, D-81730 Munich, Germany
*** Department of Electrical Engineering, Hiroshima University, 739 Higashi-Hiroshima, Japan
e-mail: uwe.feldmann@mchp.siemens.de

Abstract - Concurrent engineering aims at integrating technology, device and circuit development in parallel. We will show here an example for an EEPROM development case, which has been realized with a transient circuit simulator. A precise unified EEPROM model was developed describing all characteristics with surface potentials dependent on technological parameters. This new exact model enabled us to realize all requested goals from the first wafer run.

I. INTRODUCTION

A major request for industrial development is to minimize turn-around-time to reduce the total amount of cost. For this purpose it has been proposed to undertake concurrent engineering, synergying all contributions up to circuit design from the beginning of the development [1]. Though the technology CAD is well established with the help of process and device simulators [2], the inclusion of the circuit level is not yet well accomplished. The difficulty lies on analytical transistor models, which act as connectors between technology and circuit performance optimization. Often the models include too many model parameters, which causes the complication of the model parameter extraction [3]. One of the most serious problems is that models are not physical enough to predict technological influences on circuits correctly.

The demand for high circuit performances requires automatically a high modeling capability of the circuit simulator. Not only common MOSFET characteristics but also additional specific device features have to be included. For nonvolatile devices such as the EEPROMs studied here, the tunneling current I_{tun} into or out of the floating gate is such a feature [4,5]. This time dependent current is the essence of the EEPROM cell characteristics, and is well studied analytically [4,6,7] and numerically [8,9]. Our aim here is to show that a major issue for the successful concurrent engineering is to develop not only a reliable model for local device characteristics but also a model describing the complete EEPROM cell on the circuit level. Our newly developed model and its successful application to the EEPROM cell development will be demonstrated.

II. EEPROM FEATURES

A cross-section of the floating gate-thin oxide EEPROM cell is shown in Fig. 1. The main feature of the cell is the tunneling of charges into or out of the floating gate from or to the drain contact. To inject electrons into the floating gate for the erase operation, a positive pulse $V_{CG}(t)$ is applied on the control gate, while all other electrodes are grounded. To eject electrons out of the floating gate for the write operation, a positive pulse $V_D(t)$ is applied on the drain contact, while the control gate is grounded. The negative or positive charges in the floating gate cause a shift of the threshold voltage to higher or lower levels, which is the cause of the erase/write state.

A. Technological Aspects

Fig. 2 shows the measured threshold voltage V_{th} of the memory transistor as a function of the erase pulse duration with pulse amplitudes V_{pp} between 13V and 17V. For long durations with large V_{pp}'s a reduction of V_{th} is observed. This newly observed transient behavior cannot be explained by conventional knowledge, but by considering the charge incorporation into the interpoly dielectric ONO. The ONO structure is shown in Fig. 3. Since the bottom oxide is grown on a highly doped poly silicon, defect assisted tunneling causes charge injection into the nitride, especially at sharp grain edges, during the erase pulse with large

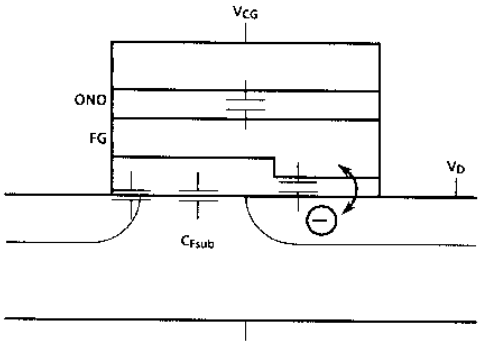

Fig. 1. Cross-section of an EEPROM cell. The essential capacitances are also indicated. Carriers flow through the tunneling oxide from the drain into the floating gate and vice versa.

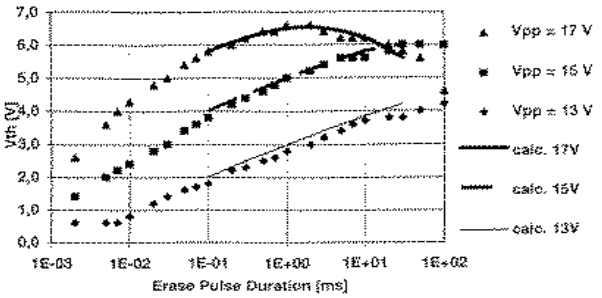

Fig. 2. Measured V_{th} as a function of the erase pulse duration for various pulse amplitudes V_{pp}. The symbols indicate measured points. The solid curves are simulated results.

amplitudes. For a long pulse duration the floating gate potential decreases due to the injected charges. This causes the increase of the electric field in the bottom oxide, which results in the charge injection into the nitride. In the nitride the charges move to the edge of the top oxide. Mostly they are trapped at the interface between the top oxide and the nitride. The V_{th} reduction is attributed to this charge movement.

The possibility of the charge loss to the drain contact through the tunneling oxide has been studied to explain the long term degradation of the EEPROM cell [10]. In our case the V_{th} reduction occurs during the pulse duration, which assists the charge injection into ONO, and not to the drain. Additionally we observed no relationship between the amount of the charge loss into ONO and the magnitude of the long-term V_{th} degradation. Therefore we have eliminated the possibility in our treatment. The charge movement in the nitride is well studied by applying a heat process to improve the charge retention of the EEPROM cell. It is shown that the bake with 250°C enables injected charges to move in the nitride [11].

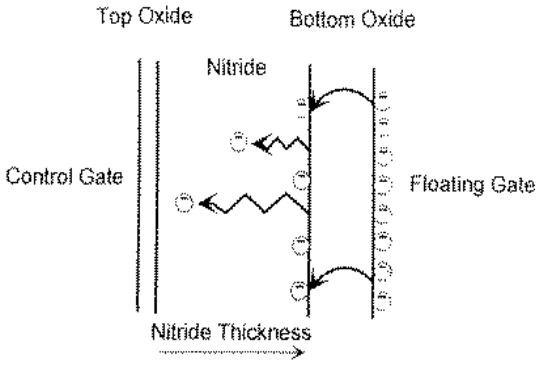

Fig. 3. After a certain pulse duration with a large amplitude charges enter into ONO. These charges contribute still to a V_{th} shift. With long pulse durations these charges move through the nitride towards the control gate, and contribute no longer to a V_{th} shift.

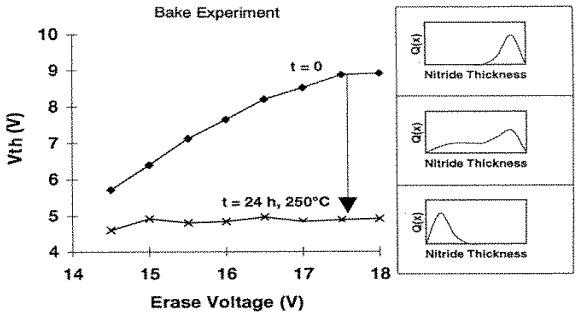

Fig. 4. Measured V_{th} of an EEPROM cell vs. erase voltage (1 ms pulse duration) before and after a 24 hours bake at 250°C.

We have performed the same experiment for 24 hours. The result is shown in Fig. 4. The V_{th} difference between before and after the bake is attributed to the movement. A nearly linear dependence of V_{th} before the bake as a function of the erase voltage is seen. The deviation from the linear dependence is caused by the charge premovement in the nitride due to the high electric field. Therefore, the evaluation of the amount of the charge movement from the static bake process is restricted only for relative low erase voltages including no premovement.

To investigate the influence of the charge movement on V_{th} qualitatively, two different nitride thicknesses T_{nit} are studied both experimentally and numerically. The bottom-oxide thicknesses are kept the same, and the top-oxide thicknesses are adjusted so that the total reduced-oxide thicknesses are the same. Fig. 5 shows the results. The measured $\Delta V_{th,b}$ values are differences between the V_{th} values with a 15V pulse of 1ms before and after the bake. The level of 15V is chosen to minimize the premovement before the bake. The numerical 2D simulation results of

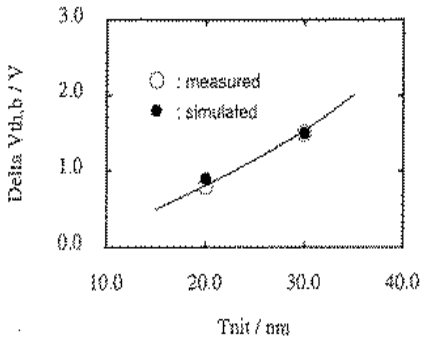

Fig. 5. Measured and simulated threshold voltage shift $\Delta V_{th,b}$ due to bake as a function of the nitride thickness. For the simulation $\Delta V_{th,b}$ is the difference between the threshold voltage with charges at the bottom oxide in the nitride and that with charges at the top oxide.

$\Delta V_{th,b}$ are differences between the V_{th} values with fixed charges in the nitride at the interface to the bottom oxide and that to the top oxide. For the simulation all inputs are kept the same except the thickness of the nitride. The fixed charge density is fitted to the measured $\Delta V_{th,b}$ value for T_{nit}=30 nm, yielding a value of $2 \cdot 10^{12}$ cm^{-2}. This charge density value is used for the T_{nit}=20 nm case to calculate $\Delta V_{th,b}$. The agreement of the simulation result with the measurement for the T_{nit}=20 nm case proves that $\Delta V_{th,b}$ is caused by the charge movement in the nitride.

Conventional specifications require 10 year data retention even at elevated temperatures. To meet this requirement we had to mitigate this effect by a slight increase of the bottom oxide thickness. However, this measure shifts the erase/write characteristics out of its optimum.

B. Circuitry Aspects

Fig. 6 shows simulated I_{tun} and the substrate current I_{sub} of an EEPROM cell. For this simulation model parameters (cf. next section) are fitted to measured I_{tun} and I_{sub}. As can be seen, I_{sub} can be much higher than I_{tun} itself. This leakage I_{sub} must be considered in memory circuit design, where the erase/write potential is provided by on chip high voltage generators. Due to the limited strength of charge pumps, the current flow during an erase or write cycle can saturate, especially when all EEPROM cells in the whole memory array are simultaneously selected. In Fig. 7 a simulated memory cell in a current saturation mode is shown. As can be seen, the increase of the floating gate potential at the beginning of an erase cycle is suppressed significantly, as indicated by an arrow, due to the I_{sub} contribution, resulting in a poor erase performance. To improve it, conventionally either the operation period is extended or the current supply is enhanced. In our case the issue will be solved by minimizing I_{sub} by optimizing the doping concentration under the tunneling oxide N_{tun}.

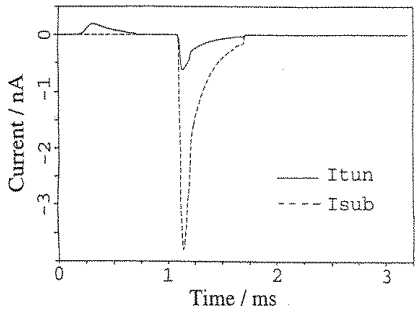

Fig. 6. Simulated transient behavior of I_{tun} and I_{sub} for an EEPROM cell at V_{CG}=0V and V_D=15V.

III. METHOD FOR CONCURRENT ENGINEERING

A. Analytical Model for the EEPROM cell

Fig. 8 shows a cross-section of the energy-band diagram of the EEPROM cell. The floating gate potential is [4,6]

$$V_{FG}(t) = \frac{C_{FC}V_{CG}(t) + C_{FD}V_D(t) - Q(t)}{C_T} \quad (1)$$
$$C_T = C_{FC} + C_{FS} + C_{FSub} + C_{FD},$$

where $Q(t)$ is the total charge stored in the EEPROM cell. All essential capacitances C are depicted in Fig. 1. The threshold voltage shift from a cell without charges in the floating gate V_{th0} is written

$$V_{th} - V_{th0} = \Delta V_{th}(t) = \frac{Q(t)}{C_{FC}}.$$

If large voltages are applied then some part of Q tunnels further into ONO, which are mostly trapped in the nitride. Fig. 2 shows that V_{th} becomes even smaller by applying large erase pulses for a long time. The long pulse duration causes the V_{FG} lowering due to the increased Q (cf. Eq. (1)). As a consequence, the potential drop across ONO increases. If the potential drop exceeds a certain limit, electrons are injected from the floating gate through the bottom oxide into the nitride. The resulting V_{th} increases further due to the additional contribution of the nitride charges, though the floating gate charge remains constant. For further pulse duration, the charges in the nitride move to the interface of the top oxide. Though these charges are in the nitride, phenomenologically they give diminished contribution to the V_{th} shift. Therefore we treat these charges as to be lost into the control gate. Consequently, the total charges stored in the floating gate are

$$Q(t) = \int_0^t [I_{tun,1}(t, E_1) + I_{tun,2}(t, E_2) - I_{PF}(t, E_2)]dt,$$

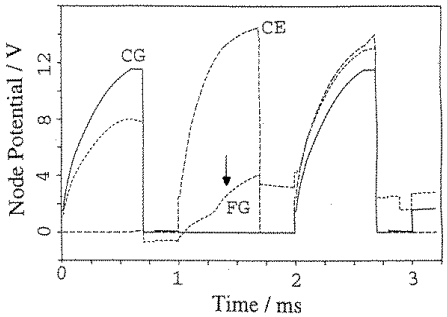

Fig. 7. Simulated transient behavior of a part of an EEPROM circuit. CG, FG, and CE are nodes on the control gate, the floating gate, and the column electrode, respectively.

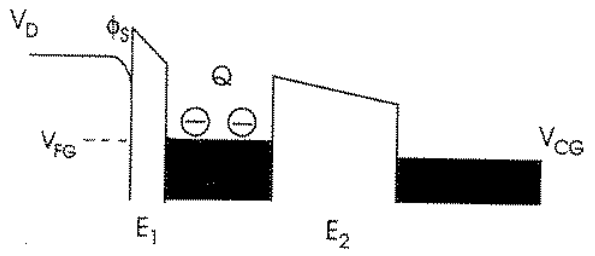

Fig. 8. Band diagram of the EEPROM cross-section.

where E is the electric field in the oxide. The tunneling current between the floating gate and the drain contact is $I_{tun,1}$ and that from floating gate into ONO is $I_{tun,2}$, and they are of a Fowler-Nordheim type which is described with two tunneling coefficients a and b [12] at an arbitrary time t

$$I_{tun}(t, E(t)) = Area \cdot a \cdot E(t)^2 e^{\frac{-b}{E(t)}}.$$

$Area$ is the tunneling area. The charge movement in the nitride is modeled by the Poole-Frenkel mechanism [13]

$$I_{PF}(t, E_2(t)) = Afloat \cdot A \cdot E_2(t) e^{-B\sqrt{E_2(t)}},$$

where A and B are treated as fitting parameters. The electric fields are

$$\begin{aligned} E_1(t) &= \frac{C_{FD}(V_{FG} - V_D - \Phi_s)}{A_{tun} \cdot \varepsilon_{ox}} \qquad (2) \\ E_2(t) &= \frac{C_{FC}(V_{CG} - V_{FG})}{A_{float} \cdot \varepsilon_{ox}}, \end{aligned}$$

where Φ_s is a surface potential depicted in Fig. 8. The areas of the floating gate and the tunneling window are defined by A_{float} and A_{tun}, respectively. The capacitances C_{FD} and C_{FC} are approximately equal to $A_{tun} \cdot \varepsilon_{ox}/T_{ox}$ and $A_{float} \cdot \varepsilon_{ox}/T_{ONO}$, respectively, where ε_{ox} is the permittivity in the oxide, T_{ox} is the tunneling oxide thickness and T_{ONO} is the effective oxide thickness of ONO.

Our intrinsic MOSFET model is based on the drift-diffusion approximation describing all transistor characteristics by surface potentials in the channel, which are computed iteratively [14]. The voltage dependence of C_{FSub} is derived automatically from the applied voltage dependence of the surface potentials, which are dependent on the technological parameters. As seen in Eq. (1), the capacitive coupling is responsible for the efficiency of the EEPROM performance. To calculate capacitances exactly, the surface potential under the tunneling oxide, Φ_s in Eq. (2), has to be calculated as well. It is computed iteratively, separately from the intrinsic MOSFET part. All above equations together with the intrinsic MOSFET model equations [14] have to be solved simultaneously.

B. Analytical Model for the Substrate Leakage Current

The undesired leakage current I_{sub} is attributed to the impact ionization by tunneling electrons from the floating gate, which is expected to occur in the space charge region [7]. These electrons have enough energy (3.2V) to produce electron-hole pairs. Most parts of these holes flow into the substrate and are observed as I_{sub}. The lateral doping profile under the tunneling oxide N_{tun} is also an important factor for the hole flow into the substrate. Our resulting I_{sub} equation is written

$$\begin{aligned} I_{sub} &= \sqrt{\frac{2\varepsilon_s \Phi_s}{qN_{tun}}} \cdot \alpha \cdot e^{-\frac{q\Phi_b}{kT}} \cdot I_{tun} \\ \alpha &= \alpha_0 \cdot e^{-(\frac{E_{crit}}{E})^\beta}, \end{aligned}$$

where α is the impact ionization rate, and α_0 is the maximum value of approximately $1 \cdot 10^5/cm$ [15]. The potential barrier Φ_b is usually around 0.2V, if the lateral diffusion of N_{tun} is far enough from the edge of the tunneling oxide [16]. The field E is given in Eq. (2) and its critical value E_{crit} is treated as a fitting parameter, and β is fixed to be one. Since our I_{sub} is observed when I_{tun} is detectable, the well known band-to-band tunneling [17,18] cannot be expected, and is neglected in our treatment.

C. Equivalent Circuit

Concurrent engineering in the EEPROM development requires a new equivalent circuit including all modeled parts. The circuit diagram of a cell is shown in Fig. 9. In addition to the conventional EEPROM equivalent circuit including a node on the floating gate, the charge injection into ONO and the movement of these charges in the nitride are explicitly introduced. The charges moved towards the control gate, which contribute no more to the EEPROM characteristics, are treated as charge loss into the control gate. This additional investigation enabled us to study the charge retention characteristics on the transient level. It is planned to handle the two extra nodes implicitly in a closed box in order to make the additional computational expense for the model evaluation not too high.

D. Parameter Extraction

To extract parameters for the EEPROM case we need a new extraction strategy and methodology, which requires special test structures. First, the extraction of the device parameters of the intrinsic MOSFET part such as the substrate doping concentration is performed for our new MOSFET model which is described in [14]. For this characterization the floating gate and the tunneling oxide window are excluded in

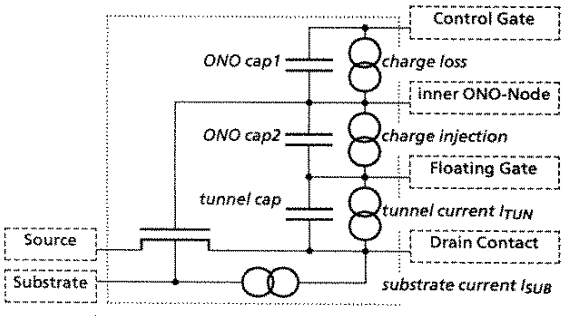

Fig. 9. Equivalent circuit diagram of the EEPROM cell. In addition to previous approaches, the charge movement in ONO is implicitly included. The charges moved towards the control gate, which do not longer contribute to the EEPROM characteristics, are treated as lost into the control gate.

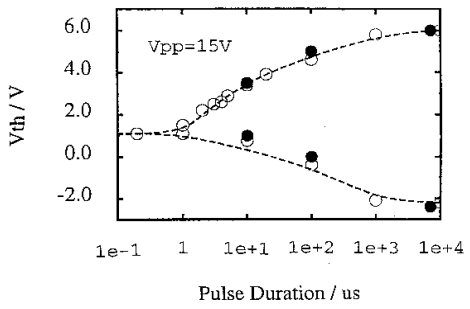

Fig. 10. Comparison of simulated V_{th} with measurements as a function of the pulse duration. Open circles are measurements and solid circles are computed results. The pulse amplitude V_{pp} (applied to the control gate for erase and to the drain for write operations) is fixed at 15V.

the test structures. With these extracted values, the tunnel coefficients a and b are extracted as a next target with the inclusion of the iteration to solve the Poisson equation under the tunneling oxide T_{ox} separately from the intrinsic channel region. This procedure is especially important to study the doping concentration influence on the tunneling efficiency. The Fowler-Nordheim tunneling coefficients into ONO are extracted from the measured V_{th} difference between before and after the bake shown in Fig. 4 with relatively small erase voltages. The Poole-Frenkel parameters are fitted to the measured V_{th} curves shown in Fig. 2. This extraction procedure can be done only with a transient circuit simulator including all modeled parts described above. The extracted parameter values are tested by comparison with the measured transient V_{th} as shown in Fig. 10.

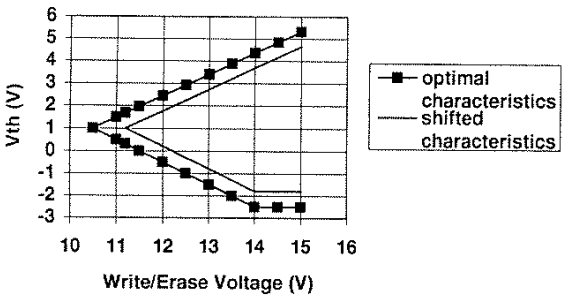

Fig. 11. V_{th} vs. write/erase pulse amplitude of an optimized EEPROM cell and of a not optimal cell with shifted characteristics due to the ONO modification (pulse duration is 1ms). The characteristics represent an average of measurements over several lots.

IV. RESULT

Our circuit simulator solves all equations described with Φ_s simultaneously. The simulation result of the charge retention is depicted in Fig. 2 together with measurements. Simulation predicts that the bottom oxide thickness has to be increased by 15 % to suppress the injection into the nitride. However, as shown in Fig. 11, the increased ONO thickness results in a V_{th} shift from the optimal value determined from the circuit operation. For compensation, our transient circuit simulation result suggests a clear advantage of the T_{ox} reduction. Fig. 12 shows the charge retention results after the optimization. For the 15V erase pulse, which is the used operation voltage in our case, no charge loss into ONO is observed. Even for 17V, the loss is drastically suppressed.

Fig. 7 shows a simulation result of a part of a memory chip in the current saturation mode. The increase of the floating gate potential at the beginning of an erasing cycle became more flat, resulting in a weaker erasing performance. During a write operation this effect is more enhanced due to the I_{sub} current flow into the substrate. The increase of N_{tun} by a factor of two suppressed the reduction of the floating-gate potential, as shown in Fig. 13. The increase of N_{tun} decreases the substrate current and thus reduces the charge pump loading. As a result we could get enough EEPROM cell performance without any change of the circuit operation condition.

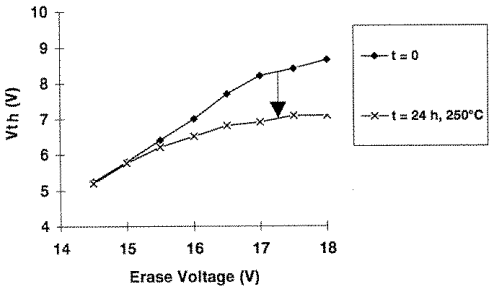

Fig. 12. Measured V_{th} of an optimized cell under the same condition as in Fig. 4. Note that a V_{th} shift by baking is observed only at high erase pulses beyond the operation voltage of 15V.

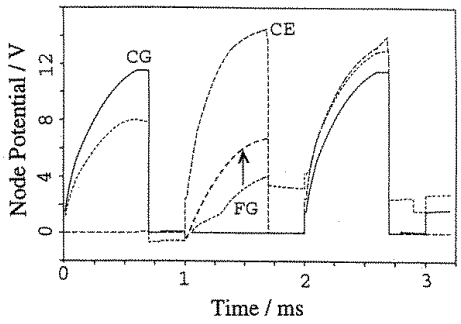

Fig. 13. Improved erasing performance by increasing the impurity concentration under the tunnel oxide.

V. Conclusion

Though numerous approaches have been investigated to characterize the retention properties, up to now no complete transient treatment of the EEPROM cell has been proposed. We have shown here for the first time the directly measured transient charge loss into the nitride in ONO. During long erase pulses these charges move to the interface between the nitride and the top oxide in ONO. This transient charge loss is modeled by three steps. The parameter extraction for the model is performed on the transient level. Together with the intrinsic transistor model part, the model was applied to optimize the EEPROM performance. It has to be emphasized that this unified optimization is possible only on the transient level with a circuit simulator due to the transient characteristics of the charge loss, which is a key optimization task.

REFERENCES

[1] S. H. Masood and B. S. Lim, "Concurrent Ingelligent Rapid Prototyping Environment," J.Intelligent Manufacturing, vol. 6, pp. 291-310, May 1995.

[2] D. M. H. Walker, J. K. Kibarian, Ch. S. Kellen, and A. J. Strojwas, "A TCAD Framework for Development and Manufacturing," Technology CAD Systems, Springer, Wien, New York, pp. 83-111, 1993.

[3] M. Kondo, T. Morie, H. Onodera, and K. Tamaru, "Model-Adaptable Parameter Extraction System for MOSFET models," IEICE Trans. Fundament. Electron. Comm. Computer Science, vol. E78A, pp. 569-572, May 1995.

[4] H. Iizuka, F. Masuoka, T. Sato, and M. Ishikawa, "Electrically Alterable Avalanche-Injection-Type MOS Read Only Memory with Stacked-Gate Structure," IEEE Trans. Elect. Devices, vol. ED23, pp. 379-387, Apr. 1976.

[5] M. -B. Chang, K. -M. Chang, C. Kuo, and S. K. Chieng, Solid-State Electron., vol. 35, pp.1521-1528, Oct. 1992.

[6] A. Bhattacharyya, "Modeling of Write/Erase and Charge Retention Characteristics of Floating Gate EEPROM Devices," Solid-State Electron., vol. 27, p p. 899-906, Oct. 1984.

[7] A. Kolodny, S. T. K. Nieh, B. Eitan, and J. Shappir, "Analysis and Modeling of Floating-Gate EEPROM Cells," IEEE Trans. Elect. Devices, vol. ED33, pp.835-844, June 1986.

[8] S. O. Kong and C. Y. Kwok, "Simulation of the program operations of the FLOTOX EEPROM," Solid-State Electron., vol. 37, pp.1949-1960, Dec. 1994.

[9] A. Concannon, F. Piccinini, A. Mathewson, and C. Lombardi, "The Numerical Simulation of Substrate and Gate Currents in MOS and EEPROMs," Tech. Dig. IEDM, pp. 289-292, 1995.

[10] C. Papadas, G. Pananakakis, G. Ghibaudo, C. Riva, F. Pio, and P. Ghezzi , "Modeling of the Intrinsic Retention Characteristics of FLOTOX EEPROM Cells under Elevated Temperature Conditions," IEEE Trans. Elect. Devices, vol. 42, pp. 678-682, Apr. 1995.

[11] C.-S. Pan, K. J. Wu, P. P. Freiberger, A. Chatterjee, and G. Sery, "A Scaling Methodology for Oxide-Nitride-Oxide Interpoly Dielectric for EEPROM Applications," IEEE Trans. Elect. Devices, vol. 37, pp. 1439-1443, June 1990.

[12] M. Lenzlinger and H. Snow, "Fowler-Nordheim Tunneling into Thermally Grown SiO2," J. Appl. Phys., vol. 4, pp. 278, 1969.

[13] C.-S. Pan, K. Wu, D. Chin, G, Sery, and J. Kiely, "High-Temperature Charge Loss Mechanism in a Floating-gate EEPROM with an Oxide-Nitride-Oxide (ONO) Interpoly Stacked Dielectric," IEEE Elect. Device Lett., vol. 12, pp. 506-509, Sep. 1991.

[14] M. Miura-Mattausch, U. Feldmann, A. Rahm, M. Bollu, and D. Savignac, "Unified Complete MOSFET Model for Analysis of Digital and Analog Circuits," IEEE Trans. Computer-Aided Design, vol. CAD15, pp. 1-7, Jan. 1996.

[15] S. M. Sze, "Physics of semiconductor devices," second ed., Wiley 1981.

[16] R. Kakoschke, JESSI report, 1996.

[17] C. Chang and J. Lien, "Corner-Field Induced Drain Leakage in Thin Oxide MOSFETs," Tech. Dig. IEDM, pp. 714-7171, 1987.

[18] T. Y. Chan, J. Chen, P. K. Ko, and C. Hu, "The Impact of Gate-Induced Drain Leakage Current on MOSFET Scaling," Tech. Dig. IEDM, pp. 718-721, 1987.

[19] E. O. Kane, J. Phys. Chem. Solids, vol. 12, pp. 181-188, 1956.