TCAD Analysis of Gain Cell Retention Time for SRAM Applications

A. Gehring, C. Heitzinger, T. Grasser, and S. Selberherr

Institute for Microelectronics, TU Vienna, Gusshausstrasse 27–29, A-1040 Vienna, Austria
Phone: +43-1-58801/36016, Fax: +43-1-58801/36099, Email: Gehring@iue.tuwien.ac.at

Abstract

We present simulations of a recently published SRAM memory gain cell consisting of two transistors and one MOS capacitor, representing an alternative to conventional six transistor SRAMs. Inverse modeling is used to fit a given device characteristic to measurement data. To account for de-charging due to tunneling, we use a simple, non-local tunneling model and calibrate it with data from literature. By optimization, we find values for the contact voltages in the off-region at which the retention time is a maximum.

1 Introduction

To meet the demand for fast, nonvolatile memory, a further increase in the density of SRAM cells is inevitable. However, common SRAM technology still relies on the conventional six transistor cell, where downsizing is difficult. Recent papers present new approaches to increase the packing density of SRAM cells using advanced cell layouts. An example is found in [1] where a MOS capacitance is used as storage node and two access transistors serve for independent read and write operations. The schematics of such a cell is shown in Fig. 1. The contacts are denoted by WWL (write word line), WBL (write bit line), RBL (read bit line) and RWL (read word line), respectively. While it is not a real SRAM circuit because of the volatility of the charge on the storage node, it offers the possibility of non-destructive read out due to capacitive coupling of the read transistor: when the storage node is charged, a positive voltage at the RWL contact suffices to open the read transistor, leading to a high current at the sense contact RBL. The sensing current is thus delivered by the RWL contact and does not reduce the charge on the storage node. The cell can be fabricated with standard process steps and consumes much less die area as compared to a six transistor SRAM.

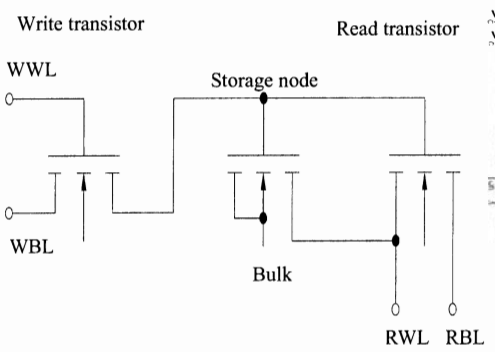

Fig. 1: Schematic of the proposed gain cell.

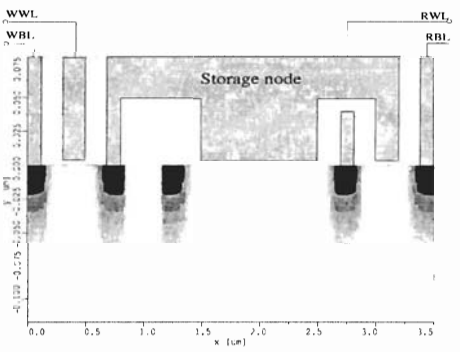

Fig. 2: Doping profile of the memory cell.

For device simulation a complete three-dimensional doping profile of the cell is not necessary. Using standard process simulation tools, single MOSFET device profiles can be used for the creation of the doping profile of the whole cell. The two-dimensional doping profile of the simulated cell is shown in Fig. 2. The doping profile was produced using the doping profile of a standard 0.2μm gate length n-type MOSFET and the process simulation tool prost2d . For such a memory cell two effects can be identified to affect the retention time: the leakage current through the access transistors and the gate tunneling current through the large gate area of the storage node.

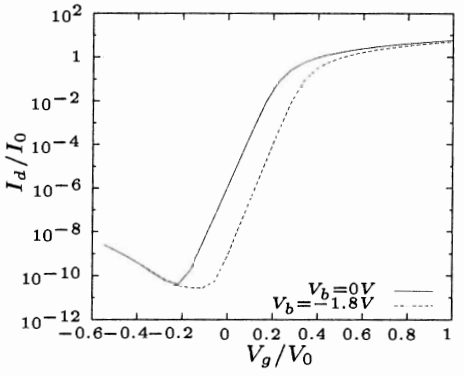

Fig. 3: Transfer chracteristics of the NMOS device.

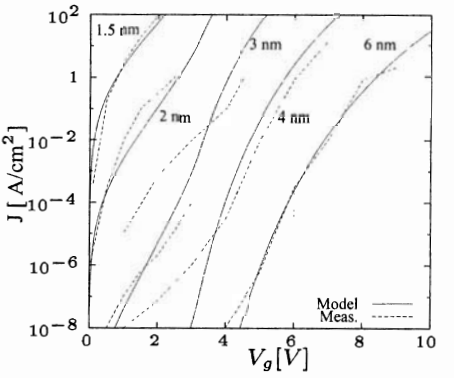

Fig. 4: Gate Current density for different gate oxide thicknesses.

2 Inverse Modeling

We used the optimization tool SIESTA [2] and the device simulator MINIMOS-NT [3] to fit the simulation models for measured transfer characteristics at different bulk biases of the NMOS device. We chose the following parameters for inverse modeling: the bulk doping for accurate modeling of the bulk voltage induced shift of the threshold voltage, the band-to-band tunneling and Shockley-Read-Hall parameters to model the drain current increase for negative gate voltages, and the work function difference. The transfer characteristics is shown in Fig. 3 for different bulk voltages, the obtained agreement to measurement data is perfect. We tried several optimization schemes (genetic optimization, gradient based optimization and simulated annealing) and achieved the best fit using simulated annealing.

3 Modeling Tunneling Currents

Due to the large size of the MOS capacitor, the storage node is de-charged by gate tunneling currents. To give some order-of-magnitude estimations, we made use of a simple non-local, electric-field based tunneling model. For low gate voltages, the SiO_2 barrier is of trapezoidal shape. Using the standard WKB approximation, the tunneling current density evaluates as

$$J = \frac{q^3 m_{ox}}{8\pi h m_0 \Phi_b} \left(\frac{V_{ox}}{t_{ox}}\right)^2 \cdot \exp\left[-\frac{4t_{ox}\sqrt{2m_{ox}}}{3qV_{ox}\hbar}\left(\Phi_b^{3/2} - (\Phi_b - qV_{ox})^{3/2}\right)\right] \quad (1)$$

where V_{ox} is the oxide voltage, t_{ox} the gate oxide thickness and Φ_b the barrier height. The other symbols have their usual meanings. The derivation of this equation can be found in [4], a similar equation is used in [5].

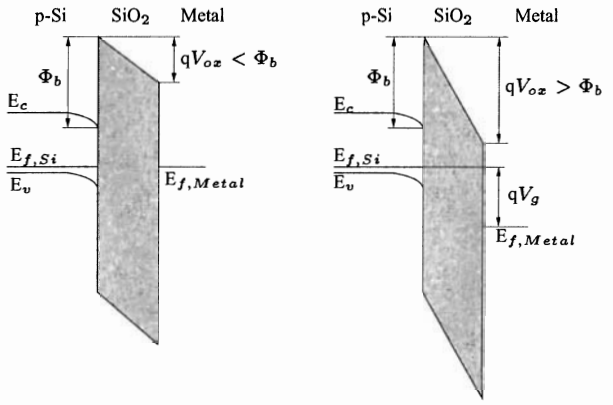

Fig. 5: Tunneling barrier for the case of direct tunneling (left) and Fowler-Nordheim tunneling (right).

As shown in Fig. 3, (1) applies for $qV_{ox} < \Phi_b >$. Otherwise, the barrier is triangular, giving rise to Fowler-Nordheim tunneling causing a gate current density of

$$J = \frac{q^3 m_{ox}}{8\pi h m_0 \Phi_b} \left(\frac{V_{ox}}{t_{ox}}\right)^2 \cdot \exp\left(-\frac{4\sqrt{2m_{ox}}}{3qF_{ox}\hbar}\Phi_b^{3/2}\right) \tag{2}$$

which is the expression usually found in literature. We used literature data to calibrate the model, with the electron mass in the oxide m_{ox} as fitting parameter and Φ_b set to 3.1 eV. In the simulation we extracted the potential values at the oxide interfaces at equidistant lateral positions beneath the gate contacts. In Fig. 4 the tunneling current densities are plotted for an oxide mass of $m_{ox} = 0.47m_0$. It shows that the model, despite of its simplicity, is accurate enough to estimate the order of magnitude of the tunneling current density over a wide range of gate voltages. The measurement values were taken from [5], [6], and [7].

4 Contact Voltage Optimization

In addition to the gate tunneling current also the voltages at the contacts in the off-region have some influence on the retention time which is caused by the leakage current of the write transistor. The leakage current of the read transistor plays a minor role since only its gate is connected to the storage node. The write transistor leakage current shows no clear dependence on the contact voltages, as shown in Fig. 6. Also the transfer characteristics in Fig. 3 justifies the assumption that there exists a minimum of the leakage current. However, since there are five independent contact voltages (WWL, WBL, RWL, RBL and the bulk contact), optimization 'by hand' becomes difficult. Thus, we again used SIESTA for optimization, employing different optimization strategies. One constraint was that all voltages in the off-region had to stay below 0.5V. For the read transistor, it turned out that positive voltages at the turned-off contacts can increase the retention time. Also, a positive bias on the bulk contact leads to a higher retention time.

5 Transient Simulation

Finally, we show results of transient simulations using the gate current model and the optimized contact voltages. In Fig. 7 the cell de-charging curves are shown for the case

of optimized and not optimized contact voltages and different gate oxide thicknesses. Only with gate oxides thinner than 2nm, a significant reduction in retention time can be seen. For optimized contact voltages, the retention time can be increased by nearly three orders of magnitude. This emphasizes the need for proper chosen contact voltages when using such devices.

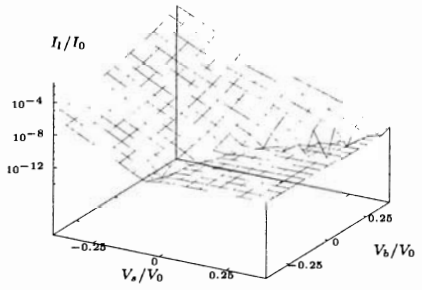

Fig. 6: Effect of source and bulk voltage on the leakage current of the write transistor.

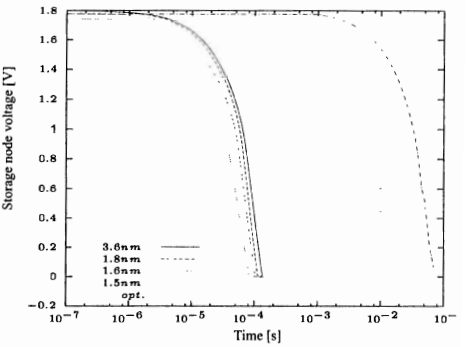

Fig. 7: De-charging curves for different oxide thicknesses (left) and optimized curve (right).

6 Conclusion

We presented simulations of a new SRAM cell consisting of two transistors and one capacitor. We used inverse modeling and optimization techniques together with rigorous device simulation to analyze the retention time of the device. We showed that, even with a very simple gate current model, measured data can be fitted to some accuracy, valuable for first estimations. However, gate current induced charge loss is crucial only for gate oxides thinner than 2nm. Of higher importance is the right choice of the contact voltages, which can increase the retention time by orders of magnitude. Without TCAD based optimization it would have been cumbersome, if possible at all, to find the right optima.

Acknowledgment

The authors acknowledge support from the "Christian Doppler Forschungsgesellschaft", Vienna, Austria.

[1] N. Ikeda, T.Terano, H. Moriya, T. Emori, and T. Kobayashi, "A Novel Logic Compatible Gain Cell with two Transistors and one Capacitor," in *Symposium on VLSI Technology Digest of Technical Papers*, pp. 168–169, 2000.

[2] C. Heitzinger and S. Selberherr, "An Extensible TCAD Optimization Framework Combining Gradient Based and Genetic Optimizers," in *Proc. SPIE International Symposium on Microelectronics and Assembly: Design, Modeling, and Simulation in Microelectronics*, (Singapore), pp. 279–289, 2000.

[3] T. Simlinger, H. Kosina, M. Rottinger, and S. Selberherr, "MINIMOS-NT: A Generic Simulator for Complex Semiconductor Devices," in *25th European Solid State Device Research Conference* (H. de Graaff and H. van Kranenburg, eds.), (Gif-sur-Yvette Cedex, France), pp. 83–86, Editions Frontieres, 1995.

[4] J. P. Shiely, *Simulation of Tunneling in MOS devices.* Dissertation, Duke University, 1999.

[5] K. F. Schuegraf, C. C. King, and C. Hu, "Ultra-thin Silicon Dioxide Leakage Current and Scaling Limit," in *Symposium on VLSI Technology Digest of Technical Papers*, pp. 18–19, 1992.

[6] N. Yang, W. K. Henson, J. R. Hauser, and J. J. Wortman, "Modeling Study of Ultrathin Gate Oxides Using Direct Tunneling Current and Capacitance-Voltage Measurements in MOS Devices," *IEEE Trans.Electron Devices*, vol. 46, July 1999.

[7] S. H. Lo, D. A. Buchanan, Y. Taur, and W. Wang, "Quantum-Mechanical Modeling of Electron Tunneling Current from the Inversion Layer of Ultra-Thin-Oxide nMOSFETs," *IEEE Trans.Electron Devices*, vol. 18, May 1997.