Simulation of a Long Term Memory Device with a Full Bandstructure Monte Carlo Approach

Christopher H. Lee, Umberto Ravaioli, *Senior Member, IEEE,* Karl Hess, *Fellow, IEEE,* Carver A. Mead, *Senior Member, IEEE,* and Paul Hasler

***Abstract*— Simulations of charging characteristics of a long term memory device, based on a floating gate structure, are presented. The analysis requires the inclusion of hot electron effects and a detailed account of the semiconductor bandstructure, because device operation is based on the injection of electrons into the gate oxide high above the silicon conduction band edge. We have developed a Monte Carlo simulator based on a full bandstructure approach which accurately accounts for the high energy tail of the electron distribution function. For practical simulation of the prototype structure, with 3.0-μm source-drain separation, the simulator is used as a post-processor on the potential profile obtained from a PISCES IIB drift-diffusion solution. The computations are in quantitative agreement with experimental results for the gate injection current, measured at fixed drain and gate biases.**

I. Introduction

We report here on Monte Carlo simulations of an analog silicon structure which is under development, for performing long term memory functions, in neural network applications. The operation of the experimental prototype is based on a single floating gate MOS structure [1]. This work is motivated by the need to design a device with a carrier injection scheme, in which the voltages are within operating limits of other devices on the same chip, using standard and widely available silicon fabrication processing [3]. A Monte Carlo approach, including the full bandstructure, has been adopted for the simulations to gain a better understanding of the hot electron, high energy tails responsible for gate currents, and to provide guidelines for future improvements. The most common forms of long term nonvolatile storage devices use charge storage on a floating gate embedded in silicon dioxide. One way to inject charge in a controllable way is to use a second electrode (select gate) in a stacked or split gate arrangement with the floating gate [2]. Large voltages are normally necessary for programming and erasing the floating gate. The structure under investigation is built similarly to an n-type channel MOS device, with source and drain n^+ contacts on a p-type silicon substrate. However, one single polysilicon floating gate is embedded in the oxide. The gate is shorter than the channel and is aligned with the edge of the drain contact, being considerably removed from the source. A schematic diagram of the structure used in the simulations is shown in the inset of Fig. 1. The prototype device has a 3.0 μm channel and a 1.0 μm gate electrode. The lateral dimension of the device is 6.0 μm. The p-type substrate, and the n-type source and drain contacts, respectively, form the base, emitter, and collector of a bipolar transistor. Conventional and well-known CMOS processes may be used to fabricate the structure.

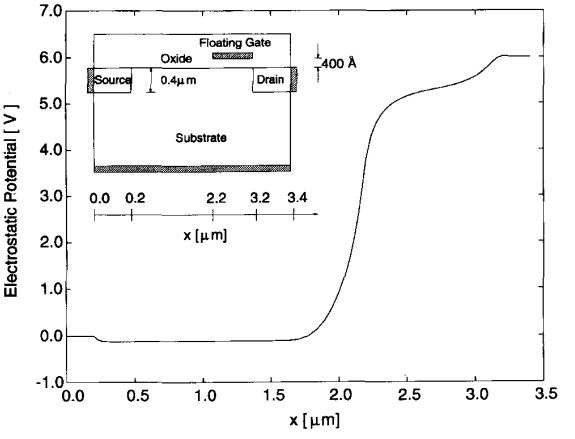

Fig. 1. Example of electrostatic potential distribution along the channel of the device with $V_D = 6.0$ V, $V_{fg} = 7.0$ V, and $V_B = 0.7$ V. The inset shows the schematic diagram of the simulated structure.

The n-type source (emitter) contact is forward biased with respect to the p-type region, acting as a source of minority carriers into the substrate (base). The positive drain voltage can accelerate the minority electrons towards the oxide interface, gaining enough energy to surmount the oxide barrier and then reach the floating gate. In order to function as a useful learning device, the charge injection rate into the floating gate should be a nonlinear function of two input variables, in this case the drain voltage and the voltage established on the gate. The gate charging behavior of the device was measured by holding gate and drain at specified voltages. The measured gate currents are shown in Figs. 2 and 3. The charging current grows exponentially at lower drain voltages for a fixed gate bias, but saturates abruptly when the drain voltage exceeds the oxide barrier potential of about 3.0 V (Fig. 2). The charging current, instead, continues to increase as the floating gate voltage is increased above the barrier energy (Fig. 3) at a fixed drain

Manuscript received January 31, 1995; revised April 26, 1995. This work was supported by the Joint Services Electronics Program, Grant N00014-90-J-1270.

C. H. Lee, U. Ravaioli, and K. Hess are with the Coordinated Science Laboratory and the Beckman Institute, University of Illinois, Urbana-Champaign, Urbana, IL 61801 USA.

C. Mead and P. Hasler are with the Department of Computer Science, California Institute of Technology, Pasadena, CA 91125 USA.

IEEE Log Number 9413026.

0741-3106/95$04.00 © 1995 IEEE

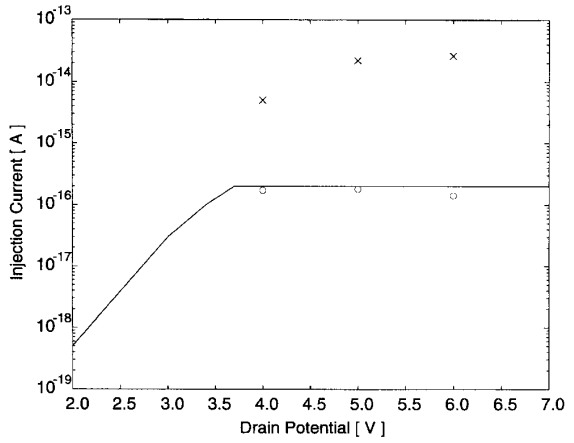

Fig. 2. Gate injection current versus drain potential. Experimental results for $V_{fg} = 5.0$ V from [1] (solid line); Monte Carlo results for $V_{fg} = 5.0$ V (o) and $V_{fg} = 7.0$ V (×).

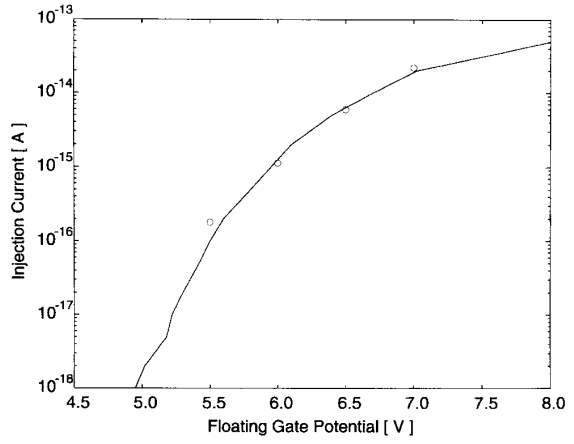

Fig. 3. Gate injection current versus floating gate potential for $V_D = 5.0$ V. Experimental results from [1] (solid line); Monte Carlo results (o).

bias. Electron injection onto the floating gate may be inhibited by reducing either the forward bias on the emitter, or the reverse bias on the drain region. These measurements of gate currents are extremely useful to establish a comparison with Monte Carlo simulations, which in turn provide an estimation of the electron energy distribution inside the structure at set bias values. Since the primary concern is long-term learning, discharge mechanisms for the floating gate are not considered here.

In order to model the device, we have implemented a simulator for MOS structures, based on a full bandstructure Monte Carlo approach, which incorporates the first two silicon conduction bands, as obtained with the empirical pseudopotential method. Complete details on the Monte Carlo approach can be found in [4]–[6]. In order to reduce the heavy computational requirements, for the large 3.0-μm channel device shown in Fig. 1 we have used the Monte Carlo program in a post-processor mode, where the potential distribution is fixed and is obtained from a drift-diffusion simulation, using PISCES IIB [8]. Checks conducted with a self-consistent version of the Monte Carlo simulator give us confidence that the results obtained in post-processor mode will agree, to within a small percentage, with fully self-consistent calculations.

We follow a procedure to evaluate the gate current, similar to that outlined by Tang and Hess [4]. Using the potential distribution from the drift-diffusion solution at the specified bias, we first inject into the source an ensemble of electrons from an equilibrium distribution. The particle energies at different locations are recorded, and particles exiting the contacts are reinjected from the source with a random energy from the equilibrium distribution, until the distribution function in the device is mapped. Then, we perform a second simulation, now injecting only the high energy tail of the distribution, starting from a suitable location of the channel before the floating gate edge. With this energy stratification approach, we can estimate the gate currents by assigning appropriate weights to the particles in the simulated tail, with respect to the total channel current detemined earlier. For oxide injection, we consider the formula for effective barrier lowering proposed by Ning [9]: $\Delta\phi_B = bF_{ox}^{1/2} + aF_{ox}^{2/3}$ [eV], where F_{ox} is the position dependent electric field in the oxide normal to the interface, with experimental parameters $a = 10^{-5}$ [cm^2 V]$^{1/3}$ and $b = 2.59 \times 10^{-4}$ [cm V]$^{1/2}$.

II. Simulation Results

The device structure used for the simulations is shown in Fig. 1. The n^+ source and drain contact regions are assumed to be uniformly doped with $N_D = 10^{18}$ cm^{-3}, and the p-type substrate, also uniform, is doped with $N_A = 10^{16}$ cm^{-3}. The source contact is grounded and the backgate contact is held at a potential $V_B = 0.7$ V, while the workfunction difference assumed between the heavily doped polysilicon gate and the substrate is approximately 0.91 V. The source-drain distance is 3.0 μm, with a 1.0-μm floating gate on the drain side. The oxide thickness below the gate is 400 Å.

The energy histogram of simulated electrons along the channel is shown in Fig. 4 for several locations under the gate. We chose the bias voltages $V_{fg} = 7.0$ V and $V_D = 6.0$ V for this example because fairly clear curves can be shown. The corresponding potential distribution along the channel, determined from PISCES IIB simulations, is shown in Fig. 1. At the left edge of the gate ($x = 2.2$ μm), the electrons are distributed over a 3.0-eV energy range, and there is a finite probability for electrons in the high energy tail to overcome the effective oxide barrier (which for this case is estimated to be about 2.71 eV, including barrier lowering [9] and energy broadening effects [4]). It is possible to recognize the signature of the bandstructure in this widely spread electron distribution. The detectable ridge in the histogram, between 1.1 V and 1.7 eV, is caused by the contribution of the L-valleys (germanium-like), and the sudden bowing between 1.7 eV and 2.3 eV reflects the sharp decrease of the total density of states which is known to occur in this energy range.

The energy spread of the histogram is rapidly compressed as the point of observation approaches locations under the

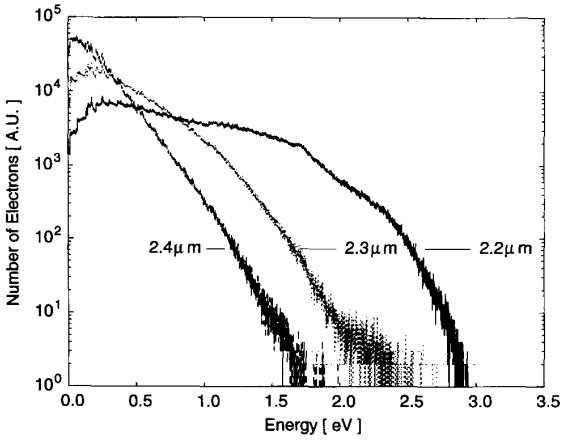

Fig. 4. Energy histogram of electrons along the channel at $x = 2.2$, 2.3, and 2.4 μm for $V_D = 6.0$ V and $V_{fg} = 7.0$ V, related to the potential distribution in Fig. 1.

gate and the high energy tail cools well below the threshold for injection into the oxide. Gate injection currents have been evaluated by using the energy stratification technique, emphasizing the statistics of the tail of the electron energy histogram. We have compared the computed results to the experimental curves available for two different bias arrangements [1]. Fig. 3 shows the measured gate current (continuous curve) obtained with a fixed floating gate voltage $V_{fg} = 5.0$ V and drain voltage V_D ranging from 2.0 to 7.0 V. The Monte Carlo results for $V_D = 4.0$, 5.0, and 6.0 V are represented by open circles and agree well with the experimental data. It should be noted that, under such operating conditions, the source current is not a sensitive function of the drain voltage because it is directed mainly toward the substrate, and is approximately constant at a value of 5.5×10^{-6} A. The drain current in this case ranges from 4.79×10^{-8} A at $V_d = 3.0$ V, to 5.05×10^{-8} A at $V_d = 6.0$ V.

To confirm the trends, we also report on the same graph Monte Carlo results for a gate voltage $V_{fg} = 7.0$ V, represented by crosses, which follow the expected behavior. We have not computed results at lower drain potentials because the gathering of gate injection statistics becomes computationally expensive, since the high energy tails become less populated.

The experimental values of gate injection current at a fixed voltage drain potential $V_D = 5.0$ V, with V_{fg} ranging between 5.0 and 8.0 V, are shown in Fig. 3. The Monte Carlo data (open circles) at $V_{fg} = 5.5$, 6.0, 6.5, and 7.0 V are again in excellent agreement. The last data point ($V_{fg} = 7.0$ V and $V_D = 5.0$ V) agrees with the Monte Carlo curve (crosses) in Fig. 2, for which a specific experimental curve was not available.

The values for the deformation potentials included in the scattering rates are those which have been used in previous applications [4]. The good agreement with the experimental results is very encouraging because it indicates that this Monte Carlo approach has sufficient predictive capabilities in resolving rare events which involve the high energy tail of the electron distribution. The calculations, conducted using the Monte Carlo model as a post-processor on a potential profile determined with a drift-diffusion approach, have been carried out on a Hewlett–Packard 700 series workstation. The full set of results required several days of computation; therefore, the approach is relatively practical, despite the computational complexity of the full band Monte Carlo method.

ACKNOWLEDGMENT

The authors thank P. D. Yoder for helpful discussions on Monte Carlo simulation.

REFERENCES

[1] J. D. W. Anderson and C. A. Mead, "MOS device for long-term learning," U.S. Patent 4,953,928, Sept. 4, 1990.
[2] *Nonvolatile Semiconductor Memories: Technology, Design, and Application,* C. Hu, Ed. New York: IEEE Press, 1991.
[3] C. A. Mead, *Analog VLSI and Neural Systems.* Reading, MA: Addison-Wesley, 1989.
[4] J. Y. Tang and K. Hess, "Theory of hot electron emission from silicon into silicon dioxide," *J. Appl. Phys.*, vol. 54, pp. 5145–5151, 1983.
[5] Y.-C. Chang, D. Z.-Y. Ting, J. Y. Tang, and K. Hess, "Monte Carlo simulation of impact ionization in GaAs including quantum effects," *Appl. Phys. Lett.*, vol. 42, pp. 76–78, 1983.
[6] *Monte Carlo Device Simulation: Full Band And Beyond*K. Hess, Ed. Norwell, MA: Kluwer, 1991.
[7] P. D. Yoder, J. M. Higman, J. Bude, and K. Hess, "Monte Carlo simulation of hot electron transport in Si using a unified pseudopotential description of the crystal," *Semicond. Sci. Technol.*, vol. 7, pp. B357–B359, 1992.
[8] M. R. Pinto, C. S. Rafferty, H. R. Yeager, and R. W. Dutton, "PISCES-IIB," Stanford University, Stanford, CA, Stanford Electron. Lab. Tech. Rep., 1985.
[9] T. H. Ning, "Hot-electron emission from silicon into silicon dioxide," *Solid-State Electron.*, vol. 21, pp. 273–282, 1978.