VLSI DESIGN
1998, Vol. 8, Nos. (1–4), pp. 13–19
Reprints available directly from the publisher
Photocopying permitted by license only

Monte Carlo Simulations of Impact Ionization Feedback in MOSFET Structures

JEFF D. BUDE

Bell Labs, Lucent Technologies, 600 Mountain Avenue, Murray Hill, NJ 07974, USA

Although impact ionization feedback is recognized as an important current multiplication mechanism, its importance as a carrier heating mechanism has been largely overlooked. This work emphasizes the inclusion of impact ionization feedback in Monte Carlo device simulations, and its implications for carrier heating in sub-micron CMOS and EEPROM technologies.

Keywords: Impact ionization, feedback, flash EEPROM, secondaries, distribution function tails

INTRODUCTION

Monte Carlo (MC) transport simulation is a widely recognized tool capable of accurately obtaining the high energy tail of the electron energy Distribution Function (DF) responsible for impact ionization, oxide degradation and gate currents, I_G, in MOSFETs. MC simulations and measurements of these hot electron effects have shown that the oxide interface DF in MOSFETs is composed of two parts [1]: the conventional channel electron DF, and a secondary electron DF coming from Impact Ionization (II) FeedBack (FB) [2]. The channel electron DF is enhanced by Electron–Electron (EE) scattering, and at low biases is principally populated by this effect. Although the secondary electron DF, generated by the coupled impact ionization of electrons and holes is usually ignored, it can dominate the channel DF tail. The relative importance of the two depends on the channel and substrate doping, drain junction depth, oxide thickness and most importantly, the substrate bias, V_{BS}. This work emphasizes the inclusion of impact ionization feedback in Monte Carlo device simulation, and its implications for carrier heating in sub-micron MOSFETs. In addition, results from a new class of EEPROM devices based on the II FB effect will be discussed.

HEATING BY II FEEDBACK IN MOSFETs

In the following, simulations have been performed using the full-band MC transport simulator, SMC [3], as a post-processor for the device simulator PADRE [4]. Doping profiles and device geometry are computed using the 2D process simulator,

PROPHET [4], and electric fields are computed for a given bias condition by PADRE. Then, given these fields, SMC solves for the device DFs in a manner similar to [5]. For more details, see [6] and references therein. Also, in this section, in order to clarify the essential elements of the II FB process in MOSFETs, EE scattering will not be considered.

Figure 1 illustrates the general phenomenon of II feedback in an *n*MOSFET [2]. Channel electrons, e_1, are injected into the drain where they II forming low energy electron-hole pairs with current II multiplication M_1. The secondary electrons, e_2 leave through the drain while the secondary holes, h_2, diffuse to the Drain-Substrate Junction (DBJ), are heated by its fields and are injected into the substrate where they II again with multiplication M_2 forming e_3 and h_3. The h_3 holes leave through the substrate, but the e_3 electrons fall back through the DBJ and vertical gate controlled potential drops reaching the oxide interface. This process continues with e_3 ionizing leading to a series of pair productions alternating between electrons and holes (II feedback) with multiplications M_3, M_4, etc. Here, $I_B = I_S(1 + M_1 + M_1M_2 + M_1M_2 M_3 + \ldots)$. The DBJ is in breakdown when this series diverges, but in the following, the devices are not in breakdown and all the $M_i < 1$.

A $L_{CH} = 0.25$ μm device with strong II FB effects has been simulated to quantify the feedback effect.

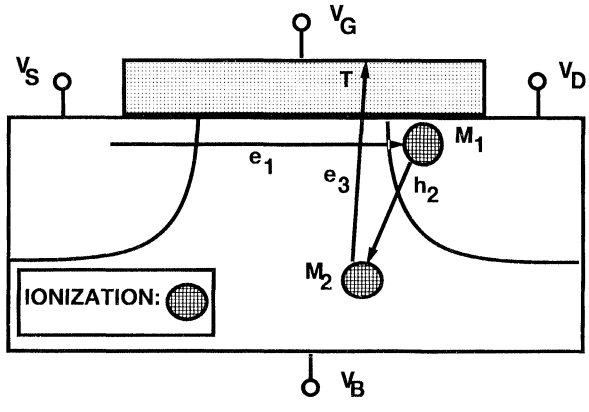

FIGURE 1 Diagram illustrating II feedback in MOSFETs.

Figure 2 shows the potential energy along the channel (from A to B in Fig. 4) and from the drain to the substrate (B to C) with $V_{GS} = 2$ V, $V_{DS} = 3$ V. Figure 3 shows the DF integrated over the device domain. The e_1 DF (no II FB) shows a rapid decay for energies above the pinch-off potential energy drop, $V_p = 2.5$ eV. There are no channel electrons above the oxide conduction band discontinuity of about $\Delta_{OX} = 3.1$ eV, but there are many at the II rate threshold of about 1.1 eV, so $I_B > 0$. The e_3 contribution is also shown in Figure 3 (II FB, full $\delta\varepsilon_{sec}$). The channel electrons ionize with $M_1 = 0.065$, initiating the FB process. The h_2 holes ionize with $M_2 = 0.02$, so that the current carried by the secondary e_3 electrons is quite small: $I_{e_3}/I_{e_1} = M_1M_2 = 0.0013$. Although the DBJ is far from breakdown, there is a broad tail extending to high energies $> \Delta_{OX}$, dominating the e_1 DF above V_p. It is sufficient to consider only the e_3 secondary electron DF since the DBJ is not in breakdown.

The energies of e_3 electrons reaching the oxide interface can be as great as $E_3^{max} = qV_{DB} + qV_{bi} + \delta_{sec}$, where $\delta\varepsilon_{sec}$ is the energy of formation of the e_3 electrons by the II of the h_2 holes, and qV_{bi} is the

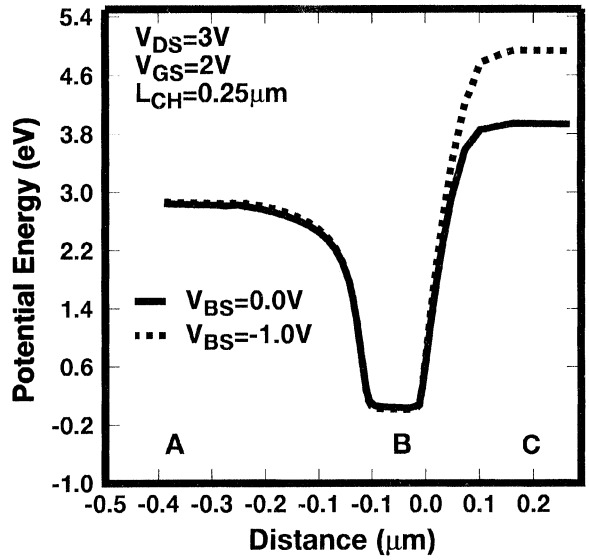

FIGURE 2 Potential energy through $L_{CH} = 0.25$ μm device of a 0.25 μm process from A–B, B–C (Fig. 1).

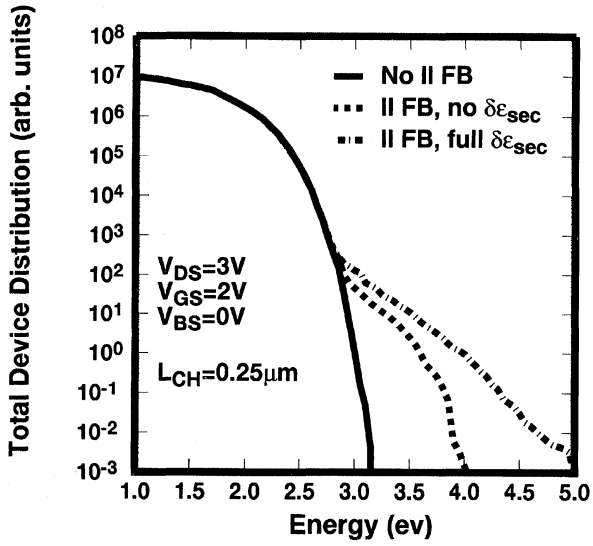

FIGURE 3 DFs integrated over the device domain for conditions and device of Figure 2.

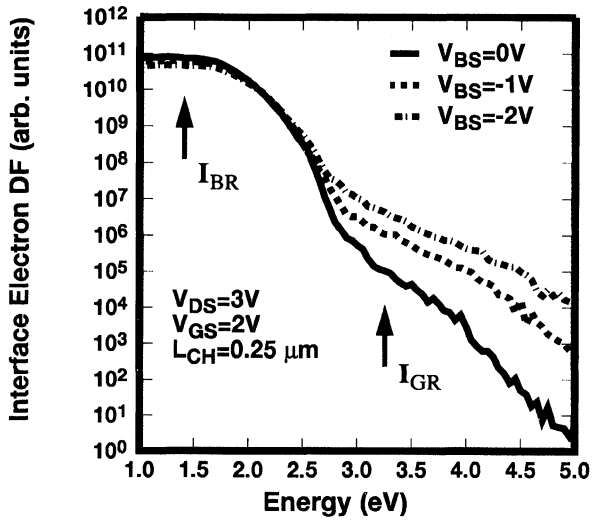

FIGURE 4 DFs integrated at the oxide interface for conditions of Figure 2. Approximate energy thresholds for substrate currents (I_{BR}) and gate currents (I_{GR}) are labelled with arrows.

built-in junction potential. Typically, $qV_{bi} \approx 0.7$ to 0.8 eV. The importance of $\delta\varepsilon_{sec}$ (a table generated during the II rate calculation [7] from the full Si band structure) is clear comparing the full $\delta\varepsilon_{sec}$ curve to that computed with $\delta\varepsilon_{sec} = 0$ (Fig. 3). The full curve extends to high energies uniformly, whereas the curve with no energy of creation begins to drop rapidly for energies near $= qV_{DB} + qV_{bi} = 4$ eV.

GATE CURRENTS BY II FEEDBACK

It is clear from above that V_{BS} should have a large effect on the tail of the DF. When a negative substrate bias, $V_{BS} < 0$, is applied, changes in the channel potential are small, but qV_{DB} changes rigidly with V_{BS} (see Fig. 2). Therefore, e_1 heating and I_B are not strongly affected, whereas e_3 heating and I_G should change exponentially. Figure 4 shows the electron DF integrated along the oxide interface for $V_{BS} = 0$, -1 and -2 V. The low energy channel DF (below 2.5 eV) changes very little (< 2 X) whereas the secondary tails responsible for I_G change exponentially ($\approx$ 1000 X) as V_{BS} goes from 0 to -2 V. This is confirmed experimentally in Figure 5 which shows $I_{GR} = I_G/I_S$ versus $I_{BR} = I_B/I_S$ for $V_{DS} = V_{GS}$ at various V_{BS}. For a given value of I_{BR} (which represents an equivalent channel heating) I_{GR} increases exponentially as V_{BS} is made more negative. This result contradicts the "effective temperature models" which predict a direct correlation between I_{GR} and I_{BR} [8] based on the assumption that both currents are generated by one thermalized channel DF. Figure 5 also shows a comparison between simulated and measured I_{GR} and I_{BR}. Qualitative trends are clearly reproduced by the simulation. Considering the approximate treatment oxide injection and transport, the remarkably good quantitative agreement with experiment is more questionable.

In addition to enhancing the high energy tail, II FB also changes the spatial distribution of the hot carriers. Figure 6 shows the number of hot secondary electrons with energies above the oxide barrier (3.2 eV) as a function of position along the oxide interface (conditions of Fig. 4). As V_{BS} is made more negative, the hot carriers spread out into the channel, enhancing L_G when V_{GS} is low. For I_G generation, the fields in the oxide at the

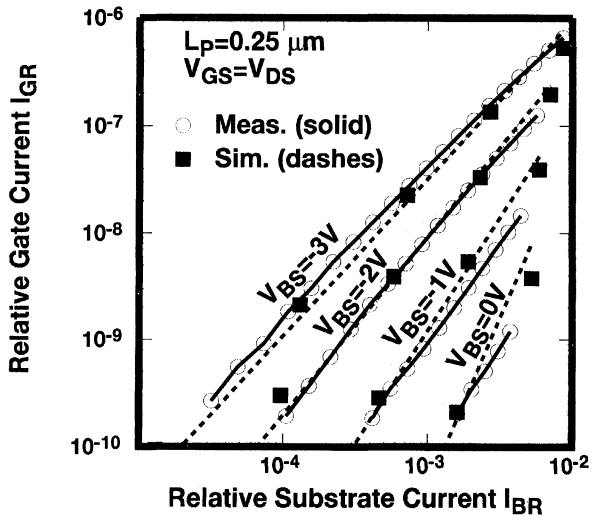

FIGURE 5 Measured and simulated I_{GR} versus I_{BR} for different V_{BS} on a $L_{CH}=0.25\,\mu m$ device. Squares, 0.5 V steps and circles, 0.1 V steps down from $V_{DS} = V_{GS} = 4\,V$.

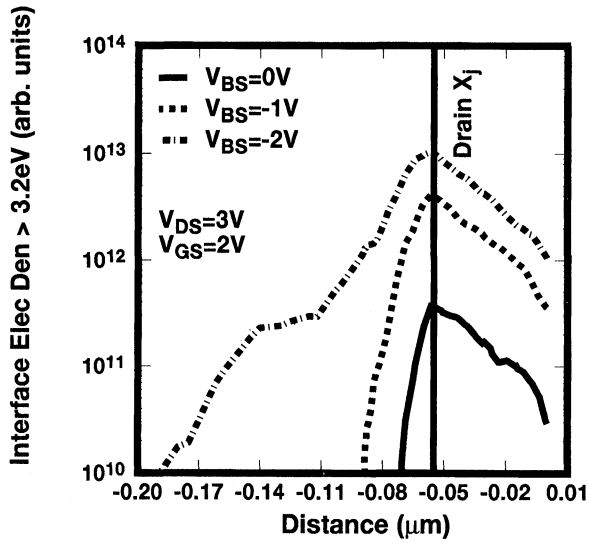

FIGURE 6 DFs integrated at the oxide interface for conditions of Figure 2.

point of injection must accelerate the electrons to the gate electrode. When $V_{GD} < 0$, accelerating fields only exist in the channel back towards the source. The secondaries, coming from the substrate, provide oxide injection at these points. This figure suggests that I_G enhancement by V_{BS} is stronger the smaller V_{GS} is compared to V_{DS}. This is confirmed in Figure 7 which shows measured gate injection efficiency, I_{GR} versus V_{GS} as a function of V_{BS} for a 0.25 μm device [9]. Note also that for $V_{BS} = -3\,V$ there is strong gate injection for V_{GS} all the way down to the threshold voltage of the device, 1.3 V. For V_{GS}, 1.3 V, I_G will go to zero with the sub-threshold drain current.

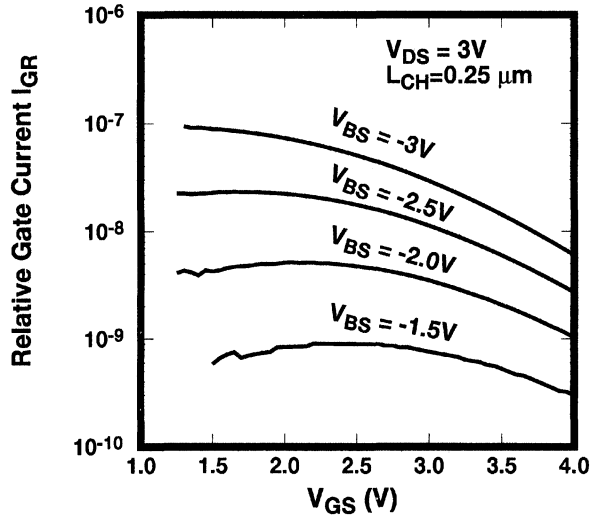

FIGURE 7 Measured I_{GR} for different V_{BS} as a function of V_{GS} on a $L_{CH}=0.25\,\mu m$ device.

TWO-PART DISTRIBUTION FUNCTIONS

As noted in the introduction, EE scattering can extend the channel DF to higher energies than V_p and V_{DS} [1, 10–11]. This EE enhanced channel DF will compete with the II FB at high energies. The relative importance of the two depends primarily on M_2 and T, factors which are controlled primarily by the sharpness of the DBJ junction and the strength of the gate field. M_2 and T become large for shallow junction devices with high channel or junction doping, and as shown above, negative V_{BS}. One easy way to determine the importance of the two is through sensitivity to V_{BS}. The channel DF is fairly independent of V_{BS},

even when EE scattering is included, but the II FB tails are strongly affected by it. In general, for a hot electron effect with a threshold energy of E_{TH}, there will always exist a V^*_{BS} such that for $V_{BS} < V^*_{BS}$, the secondaries will dominate the channel DF. For I_G, E_{TH} is 3.2 eV for points on the interface where gate fields are favorable for injection. When $V_{BS} < V^*_{BS}$, dI_G/dV_{BS} will become large. In devices of older technology generations in which II FB effect tend to be weak, $|V^*_{BS}|$ can be as high as -7 V [1]. Figure 8 shows I_G versus V_{BS} measured on two devices. The curves have been normalized by the same factor for both devices. DEV1 is a 0.25 μm device with strong FB effects similar to that of Figure 3, and DEV2 is a $L_{CH} = 0.32$ μm device optimized for high I_G (see below). The measured data was obtained by measuring floating-gate V_{TH} shifts, and the simulations were performed with both II FB and EE scattering. In DEV1, the FB DF begins to dominate I_G for $V_{BS} < -0.5$ V as shown by both measurement and simulation. In DEV2, the FB DF controls I_G even at $V_{BS} = 0$ V. Oxide DFs for DEV1 in Figure 8 are shown in Figure 9.

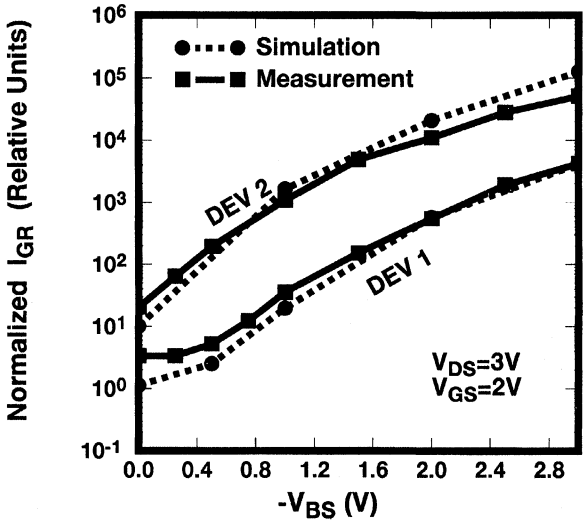

FIGURE 8 Measured and simulated I_G for two different devices: DEV1, similar to that in Figure 2, and DEV2 a $L_{CH} = 0.32$ μm device optimized for high I_G.

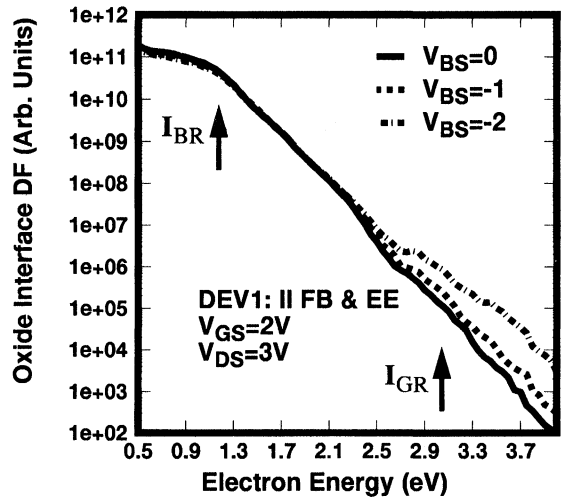

FIGURE 9 Simulated interface DFs for different V_{BS} on DEV1 for Figure 7 including II FB and EE scattering. Approximate I_B and I_G thresholds are labelled as in Figure 4.

THE CHISEL FLASH EPROM CELL

Nonvolatile memories (Flash EPROMs) use hot carrier gate injection for programming. Previous devices have relied on channel hot electron injection for programming which requires high V_{GS} (10 V or more) and V_{DS} (5–6 V) to achieve adequate programming times of 100 μs or faster. MC simulations and measurements like those shown above suggested that the II FB effect could be utilized to greatly improve programming performance for low V_{DS} and V_{GS}, leading to a lower power, better scaled EPROM cell. Based upon the physical understanding provided by these simulations, Flash EPROM devices and a writing methodology were designed to optimize the II FB process to produce a CHISEL (channel initiated secondary electron) EPROM cell [9].

Simulation optimization of the cell concentrated on the following: heavily doped junctions increase substrate current, hole multiplication, M_3 and e_3 electron transmission to the gate, T; shallow junctions enhance M_1 and T; and, $V_{BS} < 0$ enhances M_2 and T. CHISEL memory devices (Fig. 10) were fabricated by forming a stacked gate on nominal 0.25 μm nMOS devices exhibiting thin oxides and

shallow junctions (tox = 6 nm, xj = 75 nm). Because this device is based on a fully scaled CMOS technology, it is ideal for scaling into the deep submicron regime. Boron halos were added to increase junction doping, hence, M_2 and T. Most importantly, a $V_{BS} < 0$ writing scheme was adopted.

Figure 11 shows CHISEL cell programming transients for $V_{DS} = -V_{BS}$ = 2.5 and 3 V and a control gate bias of 4 V; efficient programming for times less than 100 μs can be attained with biases much lower than standard Flash devices require. Note also that the cell has strong programming until it turns itself (high I_G with floating gate potentials down to the cell threshold voltage – Fig. 7). Figure 12 highlights the strong writing time dependence on V_{BS} as anticipated.

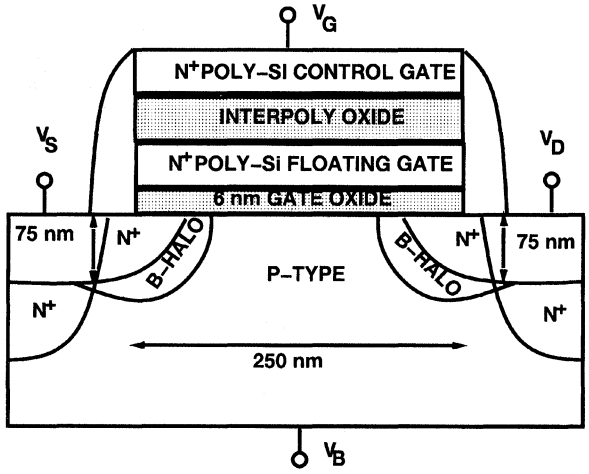

FIGURE 10 Stacked-gate CHISEL EEPROM device structure.

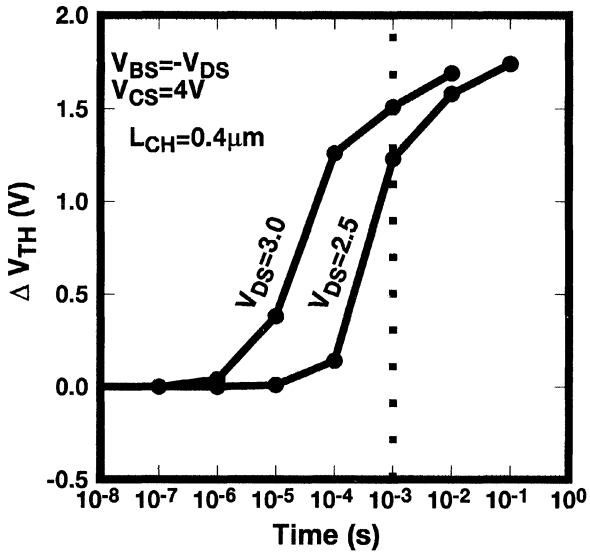

FIGURE 11 Measured programming transients for a CHISEL EEPROM cell.

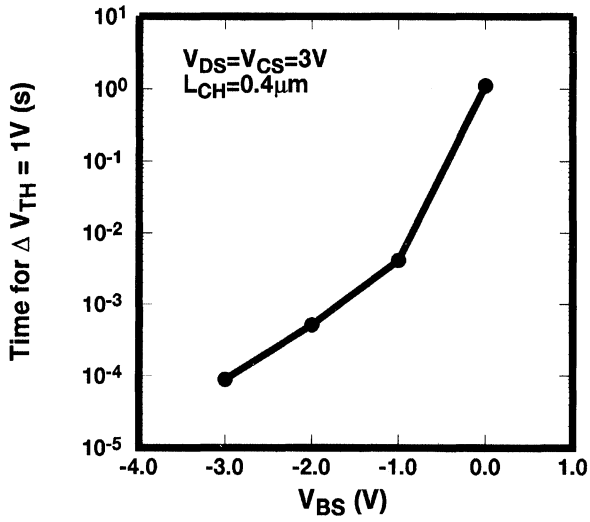

FIGURE 12 Time to reach a vth change of 1 V during programming of a CHISEL EEPROM cell.

CONCLUSIONS

Physically based MC transport simulation have clarified the importance of II FB and the two-part nature of the high energy DF tails in sub-micron MOSFETs, showing good agreement with experiment. This knowledge has led to the creation of a new class of scalable low voltage, low power EPROM devices.

References

[1] Bude, J. D. *et al.* (1996). *1996 IEDM Tech. Digest*, p. 865.
[2] Bude, J. D. (1995). *1995 Symposium on VLSI Technology*, p. 125.
[3] Bude, J. D. and Smith, R. K. (1994). In *Semiconductor Science and Technology*, HCIS 8, 9 5S, p. 840.

[4] Pinto, M. R. *et al.* (1992). *IEDM Tech. Digest*, p. 923.
[5] Higman, J. M. *et al.* (1989). *IEEE Trans. Elec. Dev.*, **ED-36**, p. 930.
[6] Bude, J. D. (1995). *Proc. of 1995 Int. Conf. on Solid State Devices and Materials*, pp. 228–230.
[7] Bude, J. D. and Mastrapasqua, M. (1995). *IEEE Elec. Dev. Lett.*, **16**, pp. 439–441.
[8] Tam, S. *et al.* (1984). *IEEE Trans. Elec. Dev.*, p. 1116.
[9] Bude, J. D. *et al.* (1995). *1995 IEDM Tech. Digest*, p. 989.
[10] Abramo, A. *et al.* (1995). *IEDM Tech. Digest*, p. 301.
[11] Fischetti, M. V. and Laux, S. E. (1995). *IEDM Tech. Digest*, p. 305.

Authors' Biography

Jeff Bude was born in St. Louis, Missouri on July 26, 1966. He received the B.S. degree with highest honors in Electrical Engineering at the University of Illinois, Urbana, in 1987. He received the M.S. degree in 1989 and the Ph.D. degree in 1992, both from the University of Illinois, Urbana. In 1992 he joined Bell Laboratories in Murray Hill, NJ as a member of the technical staff in the ULSI Technology Research Department. He has been involved in ULSI device simulation and research emphasising hot carrier effects and reliability issues.

Address: Bell Labs, Lucent Technologies, Room 2D-307B 600 Mountain Avenue, Murray Hill, NJ 07974 Phone (voice): (908) 582-3615 Phone (FAX): (908) 582-6000 e-mail: bude@lucent.com.