

# HOT-ELECTRON CURRENTS IN DEEP-SUBMICROMETER MOSFETS

J. Chung, M.-C. Jeng, G. May, P.K. Ko, and C. Hu

Department of Electrical Engineering and Computer Sciences  
University of California, Berkeley, CA 94720

## ABSTRACT

A comprehensive study of hot-electron induced substrate and gate currents in deep-submicrometer MOSFETs is presented. The  $I_{Sub}$  and  $I_{Gate}$  characteristics for devices with channel lengths as small as  $0.2\mu\text{m}$  and oxide thicknesses as thin as  $55\text{\AA}$  are examined. Implications concerning MOSFET reliability and EPROM programming are discussed. In the deep-submicrometer regime, established hot-electron concepts and models are found to be applicable; however, consideration of the finite depth of the current path and current-crowding induced weak gate control becomes much more important. With these modifications, physical analytical models for  $I_{Sub}$  and  $I_{Gate}$  are developed and verified for deep-submicrometer devices.

## INTRODUCTION

As MOSFET dimensions diminish without a corresponding reduction in power supply, hot-carrier effects continue to impose reliability limitations on device scaling [1]. The two most visible manifestations of high-field effects, substrate and gate current, have been shown to correlate with device degradation for  $n$ - [2] and  $p$ - [3] channel devices. Because  $I_{Sub}$  and  $I_{Gate}$  are sensitive functions of several parameters (such as oxide thickness, junction depth), physical understanding of these currents is useful in evaluating and comparing different device designs and structures.

This paper examines and models the substrate and gate currents in deep-submicrometer MOSFETs. Implications of device scaling on MOSFET hot-electron reliability and EPROM programming are discussed. Established longer-channel hot-electron concepts and models [4,5] are examined and found to be applicable. However, the finite depth of the current path and the effects of current-crowding induced weak gate control are additional mechanisms that must be taken into consideration. Physical analytical models for  $I_{Sub}$  and  $I_{Gate}$  are developed and verified using deep-submicrometer MOSFETs fabricated using a photoresist-ashing technique [6].

## SUBSTRATE CURRENT

**Hot-Electron Reliability Implications:** Fig. 1 shows that, as expected, peak substrate current increases with decreasing channel length and decreasing oxide thickness. It is well known that peak  $I_{Sub}$  has a close correlation with hot-electron degradation [1]. Fig. 2 shows the extrapolated maximum allowable power supply voltage to insure 10 years of lifetime as a function of  $L_{eff}$  for devices biased at peak  $I_{Sub}$ . Device lifetime is defined as the time to reach a 10% reduction in the forward linear current drive for a bias of  $V_{Gate} = 3V$  and  $V_{Drain} = 0.1V$ . For a given  $I_{Sub}$ , MOSFETs with thinner gate oxides exhibit less drain-current degradation than those with thicker oxides [7].

However, for a given drain bias, MOSFETs with thinner gate oxides also exhibit greater peak  $I_{Sub}$  than those with thicker oxides (Fig. 1). These two counteracting trends explain why the hot-electron power supply voltage in Fig. 2 is roughly independent of  $T_{ox}$ .

**Modeling  $I_{Sub}$ :** Using quasi-two-dimensional analysis, an expression for the maximum lateral electric field has been derived [1,4,8]

$$E_{max} = (V_{Drain} - V_{DSAT})/l \quad (1)$$

where  $V_{DSAT}$  is the drain saturation voltage and  $l$  is the length of the velocity-saturation region, which is a function of  $T_{ox}$  and  $x_j$ . Substrate current can be expressed as [4]

$$I_{Sub} = I_{Drain} (\alpha_i/\beta_i) (E_{max}/l) e^{-\beta_i/E_{max}} \quad (2)$$

where  $\alpha_i$  and  $\beta_i$  are the impact ionization coefficients. Typical values for  $\alpha_i$  and  $\beta_i$  are  $\alpha_i = 2.5 \times 10^6 \text{cm}^{-1}$  and  $\beta_i = 2 \times 10^6 \text{V/cm}$  [1,9].

In order to account for nonlocality of the impact ionization coefficient, modifications to Eq. 2 have been made. Based on Monte-Carlo simulations and theoretical considerations [10], it has been proposed letting  $\alpha_i = \gamma l$  where  $\gamma = 3 \times 10^{10} \text{cm}^{-2}$ . By treating nonlocal effects as a weak parameter dependence in  $\alpha_i$ , the basic functional form of the  $I_{Sub}$  model is left unchanged.

**Scaling of the Velocity-Saturation Region Length,  $l$ :** Using Eqs. 1 and 2, the parameter  $l$  can be extracted from the slope of a plot of  $I_{Sub}/[I_{Drain}(V_{Drain} - V_{DSAT})]$  versus  $1/(V_{Drain} - V_{DSAT})$  [11] as shown in Fig. 3. The relation  $l = 0.22 T_{ox}^{1/3} x_j^{1/3}$  [11,12] has been observed for long-channel/thick-oxide devices; similar dependences are also seen in theoretically derived  $l$  [4,8]. The linearity of the experimental data in Fig. 3 suggests that the simple electric field and substrate current model underlying Eqs. 1 and 2 is valid even for deep-submicrometer devices. In Fig. 3,  $l$  is observed to decrease sub-linearly with  $L_{eff}$  ( $l \propto L_{eff}^{1/5}$ ) for  $L_{eff} < 0.5\mu\text{m}$ . This behavior is probably due to the encroachment of the linear region into the velocity-saturation region as  $L_{eff}$  decreases.

In Fig. 4,  $l$  is observed to become a less sensitive function of oxide thickness ( $l \propto T_{ox}^{1/8}$ ) for  $T_{ox} < 150\text{\AA}$ . One explanation for this behavior is the non-negligible depth of the current path. In [4,8], impact ionization is assumed to occur at the surface. However, in saturation, peak impact ionization actually occurs hundreds of angstroms below the surface. Thus according to the quasi-two-dimensional model,  $l$  should scale with an "effective" oxide thickness consisting of  $T_{ox}$  and the depth of the current path. Only when  $T_{ox}$  becomes comparable to the current path depth, does  $l$  become less sensitive to the oxide thickness; this explains why the  $T_{ox}^{1/8}$  dependence is not observed in [11,12].

In Fig. 4, the separation between the groups of data points is due to differences in junction depth. Using PISCES II [13], the dependence  $l \propto x_j^{1/3}$  is found to remain unchanged even for deep-submicrometer devices. Thus, for  $L_{eff} < 0.5\mu\text{m}$  and  $T_{ox} < 150\text{\AA}$ , the following relation for  $l$  should be used

$$l = 1.7 \times 10^{-2} T_{ox}^{1/8} x_j^{1/3} L_{eff}^{1/5} \quad (3)$$

where  $l$ ,  $T_{ox}$ ,  $x_j$ , and  $L_{eff}$  are all expressed in units of  $cm$ .

**Current-Crowding Induced Weak Gate Control:** Fig. 5 shows that as  $L_{eff}$  is reduced,  $I_{Sub}$  decreases less rapidly with increasing  $V_{Gate}$ . One explanation for this behavior is current-crowding induced weak gate control (illustrated in Fig. 6). When a large amount of current (either from high gate bias, thin gate oxide, or short channel length) is forced to flow into a shallow junction, the mobile carrier concentration  $Q_{mob}$  becomes comparable to the doping concentration in a portion of the drain outside the full gate-controlled region, thus forming a weak gate-controlled region of length  $L_D$ . The high lateral electric field arising in  $L_D$  results in increased  $I_{Sub}$  [14]. As illustrated in Fig. 6, the same drain-junction technology that yields good gate-control for long-channel devices can produce weak gate-control behavior in short-channel or thin-oxide devices.

The length of the weak gate controlled region is a function of the drain doping profile, the gradient of the drain doping profile, and  $Q_{mob}$ . The increase in the maximum lateral electric field can be modeled as

$$\Delta E_{max} = \left[ \frac{L_D}{\epsilon_{si}} \right] \left[ \frac{qI_{Drain}}{x_j v_{sat}} - N_{Drain} \right] \quad (4)$$

where  $v_{sat}$  is the carrier saturation velocity,  $\epsilon_{si}$  is the silicon dielectric constant, and  $N_{Drain}$  is the drain junction doping.

**Analytical  $I_{Sub}$  Model Results:** An analytical substrate current model based on Eq. 2 and incorporating Eqs. 3 and 4 has been developed. A hot-electron based drain current model is used for  $I_{Drain}$  [15]. Fig. 7 shows that a value of  $L_D = 50\text{\AA}$  produces agreement with the experimental results. In Fig. 8, good agreement is observed between the  $I_{Sub}$  model and measured results for a wide range of  $L_{eff}$ .

#### GATE CURRENT

**EPROM Programming and Scaling Implications:** Fig. 9 displays the bias ( $V_{Gate} = V_{Drain}$ ) necessary for  $I_{Gate}/W_{eff} = 100\text{pA}$  as a function of  $L_{eff}$  for differing  $T_{ox}$ . It appears that 5V EPROM programming may require as small as  $0.4\mu\text{m}$  devices for reasonable programming speed. In addition, Fig. 9 suggests an upper bound to the power supply for normal device operation in order to prevent excessive gate hot-electron leakage current.

In Fig. 1 and 9,  $T_{ox}$  is observed to have different effects on the hot-electron currents. Reducing  $T_{ox}$  diminishes  $l$  (Eqn. 3) which enhances  $E_{max}$  (Eqn. 1). However, at high  $V_{Gate}$ , reducing  $T_{ox}$  also increases  $V_{DSAT}$  [15] which reduces  $E_{max}$  (Eqn. 1). These two counteracting mechanisms explain why peak  $I_{Sub}$ , which occurs primarily at low  $V_{Gate}$ , increases with decreasing  $T_{ox}$ , whereas  $I_{Gate}$ , measured at  $V_{Gate} = V_{Drain}$ , decreases with decreasing  $T_{ox}$ .

Fig. 10 shows that, for deep submicrometer devices,  $I_{Gate}$  does not decrease with increasing  $V_{Gate}$  as is observed in longer-channel devices. The mechanism of current-crowding induced weak gate control also explains this behavior. Because  $E_{max}$  does not decrease as rapidly with increasing  $V_{Gate}$  as in long-channel devices, and because oxide-barrier lowering [16] with increasing  $V_{Gate}$  remains unchanged,  $I_{Gate}$  increases.

**Model for  $I_{Gate}$ :** An expression for gate current based on the lucky-electron model has been developed [5]

$$I_{Gate} = F(E_{ox}) I_{Drain} e^{-\Phi_B / (q\lambda E_{max})} \quad (5)$$

where  $F(E_{ox})$  is weak function of the oxide field and  $\Phi_B$  is the effective barrier height (including barrier lowering). However,  $I_{Gate}$  has been measured at drain voltages well below  $\Phi_B$  [17]; this phenomenon has also been observed in this study. This fact would tend to suggest that an effective-temperature rather than a lucky-electron model be used. If an effective-temperature model is adopted, then a similar expression to Eq. 5 can still be obtained by making the substitution  $T_e = qE_{max}\lambda/k$  [17].

According to Eqs. 2 and 5, a correlation is expected between  $I_{Gate}/I_{Drain}$  and  $I_{Sub}/I_{Drain}$  [5]; this is illustrated in Fig. 11. The fact that the correlation still applies for deep submicrometer devices, suggests that the basic  $I_{Sub}$  and  $I_{Gate}$  models are still valid in that regime.

**Analytical  $I_{Gate}$  Model Results:** An analytical gate current model based on Eq. 5, incorporating Eqs. 3 and 4 and a hot-electron based  $I_{Drain}$  model [15], has been developed. In Fig. 12, good agreement is observed between the  $I_{Gate}$  model and measured results for a wide range of  $L_{eff}$ .

#### CONCLUSIONS

In this study, the hot-electron currents in deep-submicrometer MOSFETs have been comprehensively analyzed. With appropriate modifications, existing  $I_{Sub}$  and  $I_{Gate}$  concepts and models are found to be applicable for deep-submicrometer devices.

#### ACKNOWLEDGEMENT

This project was funded by SRC, JSEP under contract F49620-84-C-0057, Signetics/Philips, the California State MICRO program, and ISTO/SDIO administered through ONR under contract N00014-85-K-0603.

#### REFERENCES

- [1] C. Hu, et al, *IEEE Trans. Electron Devices*, ED-32, p.375, Feb. 1985.
- [2] E. Takeda, et al, *IEDM Tech. Dig.*, p. 60, 1985.
- [3] M. Brassington, et al, *Symposium on VLSI Technology Tech. Dig.*, p. 55, 1987.
- [4] P. Ko, et al, *IEDM Tech. Dig.*, p. 600, 1981.
- [5] S. Tam, et al, *IEEE Trans. Electron Devices*, ED-31, p.1116, Sep. 1984.
- [6] J. Chung, et al, *IEEE Elec. Dev. Lett.*, EDL-9, p. 186, Apr. 1988.
- [7] M. Yoshida, et al, *IEDM Tech. Dig.*, p. 254, 1985.
- [8] Y. El Mansy, et al, *IEEE Trans. Electron Devices*, ED-24, p. 254, Mar. 1977.
- [9] J. Slotboom, et al, *IEDM Tech. Dig.*, p. 494, 1987.
- [10] J. Higan, et al, *IEEE Elec. Dev. Lett.*, EDL-9, p. 399, Aug. 1988.
- [11] T. Chan, et al, *IEEE Elec. Dev. Lett.*, EDL-6, p. 551, Oct. 1985.
- [12] M. Kakumu, et al, *IEDM Tech. Dig.*, p. 399, 1986.
- [13] M. Pinto, et al, *Stanford Elec. Lab. Tech. Rep.*, Stanford Univ., Feb. 1986.
- [14] P. Ko, et al, *IEDM Tech. Dig.*, p. 292, 1986.
- [15] M. Jeng, et al, *IEDM Tech. Dig.*, p. 710, 1987.
- [16] T. Ning, et al, *IEEE Trans. Electron Devices*, ED-26, p. 346, Apr. 1979.
- [17] S. Tam, et al, *IEEE Elec. Dev. Lett.*, EDL-4, p. 249, Jul. 1983.

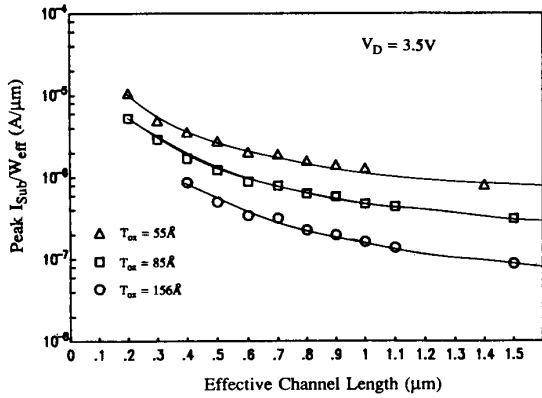


Figure 1. Peak substrate current versus  $L_{eff}$  for  $T_{ox} = 55\text{\AA}$ ,  $85\text{\AA}$ ,  $156\text{\AA}$ . The drain voltage is  $3.5\text{V}$ .

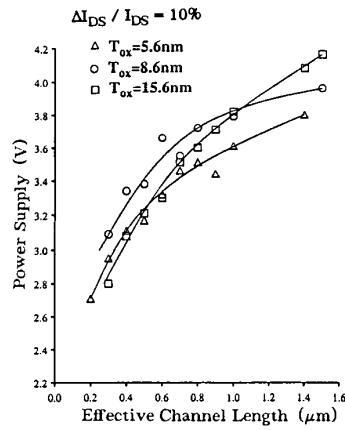


Figure 2. Voltage required to insure 10 years of device lifetime for  $T_{ox} = 56\text{\AA}$ ,  $86\text{\AA}$ ,  $156\text{\AA}$ .

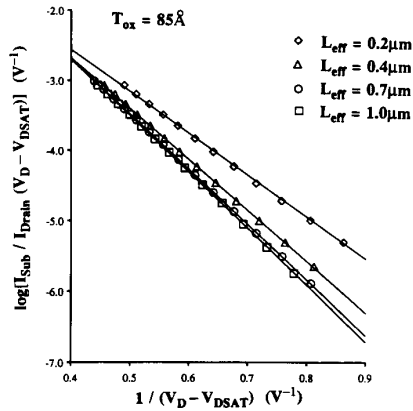


Figure 3. Plot of  $\log[I_{sub}/I_{drain}(V_D - V_{DSAT})]$  versus  $1/(V_D - V_{DSAT})$  for  $L_{eff} = 0.2\mu\text{m}$ ,  $0.4\mu\text{m}$ ,  $0.7\mu\text{m}$ ,  $1.0\mu\text{m}$ . The oxide thickness is  $85\text{\AA}$ .

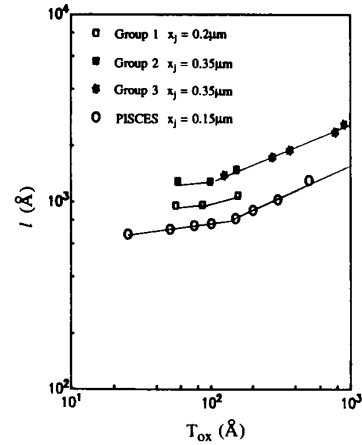


Figure 4. The velocity-saturation region length  $l$  as a function of oxide thickness. Separation between the groups of data points is due to differences in  $x_j$ .

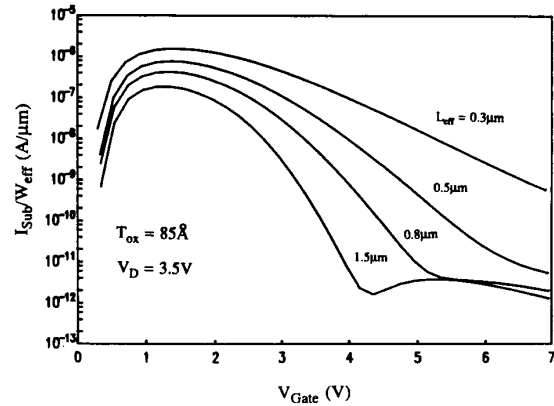


Figure 5. Substrate current versus  $V_{Gate}$  for  $L_{eff} = 0.3\mu\text{m}$ ,  $0.5\mu\text{m}$ ,  $0.8\mu\text{m}$ ,  $1.5\mu\text{m}$ . The oxide thickness is  $85\text{\AA}$ . The drain voltage is  $3.5\text{V}$ .

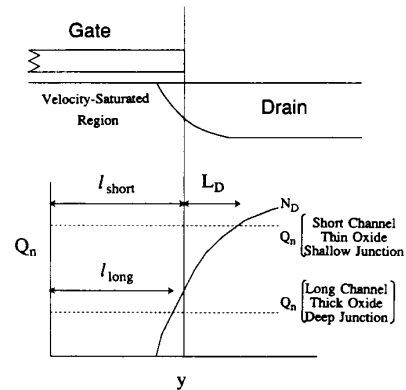


Figure 6. A physical model for the weak gate-controlled region.  $L_D$  represents the extension of the velocity-saturated region of the channel  $l$  into the drain.

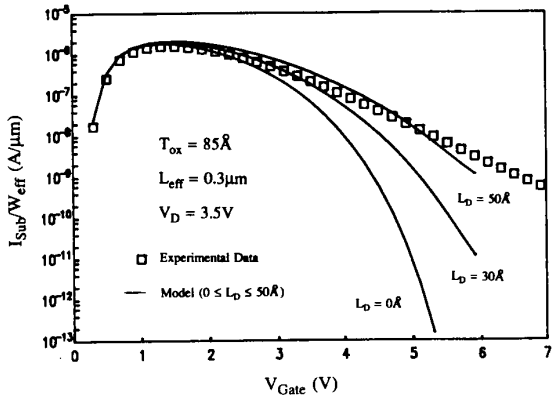


Figure 7. Substrate current versus  $V_{Gate}$  for  $T_{ox} = 85 \text{ \AA}$  and  $V_D = 3.5V$ . The model for 3 values of  $L_D$  is compared with the measured data.

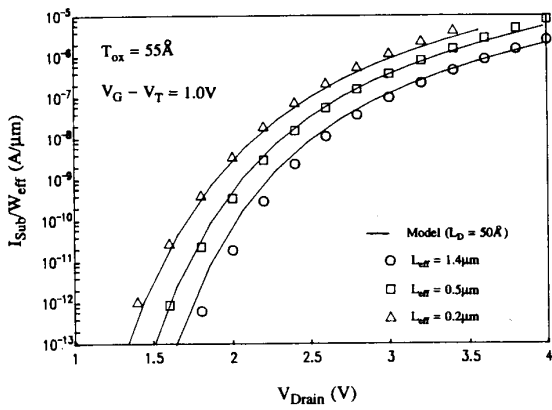


Figure 8. Substrate current versus  $V_{Drain}$  for  $L_{eff} = 0.2 \mu m, 0.5 \mu m, 1.4 \mu m$ . The oxide thickness is  $55 \text{ \AA}$ . Model and experimental results are presented.

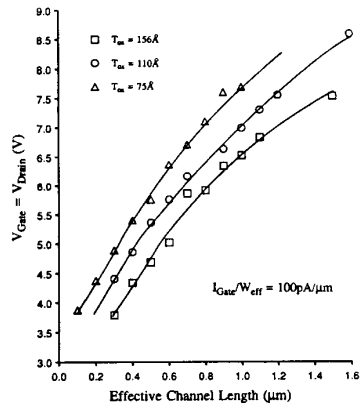


Figure 9. Voltage ( $V_{Gate} = V_{Drain}$ ) required for  $I_{Gate}/W_{eff} = 100 \text{ pA}/\mu m$  as a function of  $L_{eff}$  for  $T_{ox} = 75 \text{ \AA}, 110 \text{ \AA}, 156 \text{ \AA}$ .

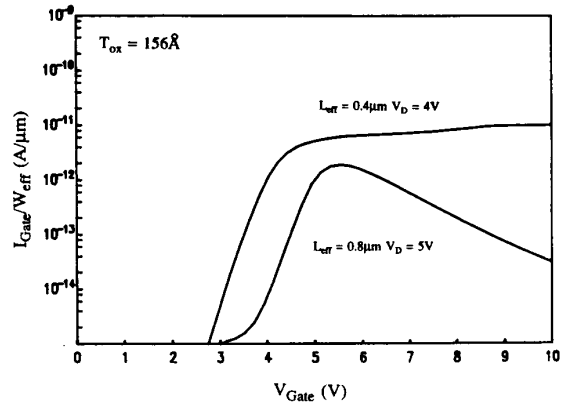


Figure 10. Gate current versus  $V_{Gate}$  for  $L_{eff} = 0.8 \mu m, V_D = 5V$  and  $L_{eff} = 0.4 \mu m, V_D = 4V$ . The oxide thickness is  $156 \text{ \AA}$ .

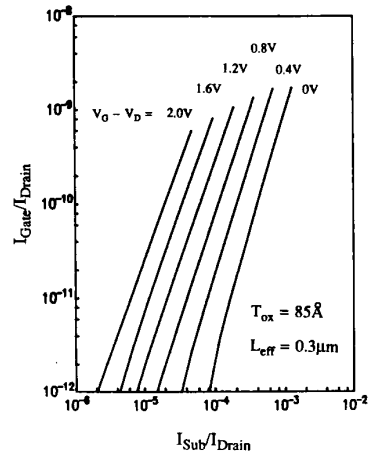


Figure 11.  $I_{Gate}/I_{Drain}$  versus  $I_{Sub}/I_{Drain}$  for constant  $V_G - V_D$ . The channel length is  $0.3 \mu m$ . The oxide thickness is  $85 \text{ \AA}$ .

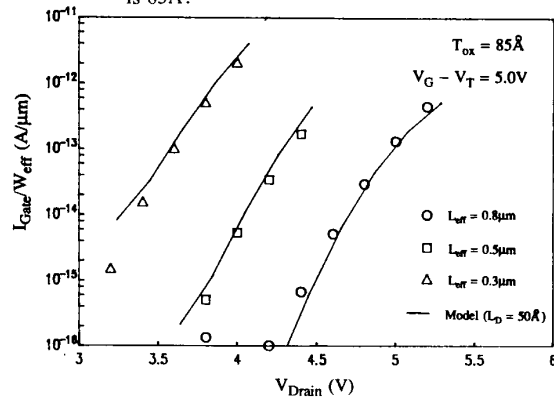


Figure 12. Gate current versus  $V_{Drain}$  for for  $L_{eff} = 0.3 \mu m, 0.5 \mu m, 0.8 \mu m$ . The oxide thickness is  $85 \text{ \AA}$ . Model and experimental results are presented.