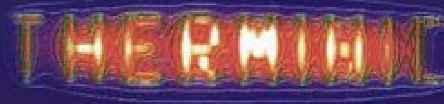

Aix-en-Provence, France, 24-26 September 2003

A NEW SIMULATION MODEL OF ELECTRO-THERMAL DEGRADATION FOR MOSFET DEVICES SUBJECTED TO HOT CARRIER INJECTION STRESS

Gunhan Kaytaz, Pavel L. Komarov, and Peter E. Raad

Mechanical Engineering Department
Southern Methodist University
Dallas, TX 75275-0337
praad@smu.edu

ABSTRACT

This work presents a new electro-thermal degradation simulation method to assess the approximate lifetime of an active MOSFET subjected to hot carrier injection stress. Data obtained from short-term stressing measurements were used to establish a model capable of predicting long-term degradation effects. The resulting model was checked against actual long-term MOSFET degradation experiments and found to be accurate for the range of oxide thickness values considered in this investigation. The resulting method can be used to guide the analysis and design of high power density integrated circuits.

1. INTRODUCTION

Electronic devices are getting smaller, faster, more efficient, and more reliable. However, reliability assessment remains paramount to predict average lifetime of a device typically made up of many integrated circuits. The assessment is usually accomplished by monitoring the performance of a device under working conditions for long periods of time, which can be costly and time consuming. Therefore, it is desirable to minimize testing duration. To that end, a reliable prediction tool capable of simulating the device life-performance relationship is essential. The majority of the work that has been done in this area [1-3] has been focused on determining the effects of degradation by the use of electrical simulation and modeling. Since temperature usually is taken to play a small part in these models, the effects of temperature levels (and variations) on device functionality are not captured [4]. But in order to better understand the effects of degradation, the thermal effects must be fully coupled with the electrical effects within a comprehensive electro-thermal simulation method, particularly since elevated temperatures contribute to the contamination of the silicon oxide barrier layer, a phenomenon known as hot carrier injection.

This work introduces a new electro-thermal degradation simulation method and suggests its use in determining the approximate lifetime of an active MOSFET subjected to hot carrier injection stress. Hot carrier injection is used to an advantage in EPROM applications, but in other MOSFET applications it is the main reason for device failure. Carriers, which gather enough energy from higher temperatures to jump the silicon dioxide barrier can get trapped inside the oxide layer, deteriorating the electrical properties of the latter, and thereby shifting device properties and causing performance degradation. In extreme cases, this process may lead eventually to the breakdown of the silicon dioxide layer [5]. It is necessary to accurately predict the change in local parameters as a result of the elevated temperature field and translate that information into a performance drop function, which leads to an approximation of useful life span of a MOSFET.

2. NUMERICAL SIMULATION

An electro-thermal simulation method has been developed, providing the ability to determine the changes in local variables and properties of an active MOSFET. The numerical approach consists of three main parts: (i) a thermal simulation to solve the energy equation in a given device domain, (ii) an electrical device simulation written specifically for MOSFET devices, and (iii) a predictive solution of the degradation in device performance.

During a simulation, the thermal and electrical simulation modules interact with each other. The local heat source term used in the thermal simulation is calculated by the electrical module, which in turn uses the temperature distribution as a basis to modify the device variables and properties. The combined effects of thermal

and electrical behaviors are used as input for the degradation module, which in turn changes physical properties of the oxide layer in order to represent the deterioration during the time duration being considered.

The thermal response is obtained by solving the two-dimensional heat transfer equation (Eq. (1a)), where the heat source, $\dot{Q}_{el}$, is calculated from the electric field, $\nabla\phi_{loc}$, and local current density, J_{loc}, (Eq. (1b)):

$$\rho c_p \frac{\partial T}{\partial t} = \nabla \cdot \left(K \nabla T\right) + \dot{Q}_{el} \quad (1a)$$

$$\dot{Q}_{el} = J_{loc} \nabla \phi_{loc} \quad (1b)$$

In all cases under investigation, MOSFETs with a gate width that is an order of magnitude bigger than the gate length were chosen to obtain high output power from the device. Therefore, the three-dimensional heat transfer equation was reduced to its two-dimensional form. Changes in thermal conductivity of the materials with temperature are included through empirical relations [6]. The energy equation is solved by a Padé-based three-point backward finite difference approximation. The three-point backward scheme is unconditionally stable and is second-order accurate in space and time.

The boundary conditions for the heat transfer equation are chosen such that the top surface of the MOSFET is adiabatic and the back contact is held at a constant temperature. In the horizontal plane, locations that are 1μm beyond the far edge of the doped regions are considered to be adiabatic (see Fig. 1).

The grid used in the calculation is a simple rectangular grid, which is fine enough to capture the characteristics of the MOSFET. The cell size in the y-direction is chosen to be small enough to include at least 10 computational cells within the thinnest layer. Most of the time, the oxide layer thickness is the determining factor in cell size. In the x-direction, the cell size is chosen to be several times larger than in the y-direction.

The initial condition for the heat transfer equation is chosen to reflect a uniform temperature field at all locations on the computational grid, i.e., at $t = 0$, $T = T_{ref}$ at every point on the MOSFET, where T_{ref} was chosen to be ambient temperature.

The electrical simulation is based on solving the local current density (Eq. (2)) and potential field (Eq. (3)) equations [7] using a variable depletion layer thickness model [8], modified to reflect the changes in temperature (Eq. (4)). The current density consists of a drift-driven term and a diffusion-driven current term (Eq. (2)). Because of the nature of the samples used in this work, the current is mainly the result of drift, rather than diffusion. Therefore, the diffusion term is neglected since it is often very small compared to the drift term.

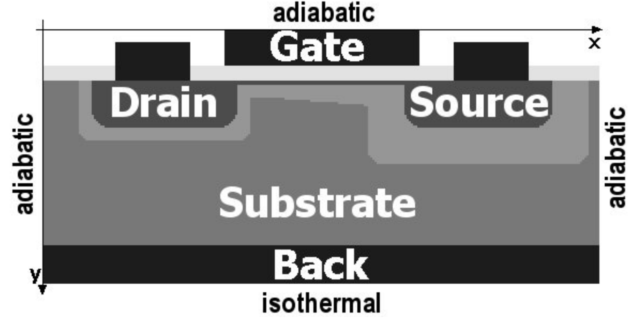

Fig. 1 Schematic of a sample MOSFET under investigation

$$\vec{J} = q\left[\underbrace{\left(\mu_n n + \mu_p p\right)\nabla V}_{drift} + \underbrace{D_n \nabla n - D_p \nabla p}_{diffusion}\right] \quad (2)$$

In Eq. (2), $\vec{J}$ denotes the current density; μ_n and μ_p are electron and hole mobility, respectively; n and p are electron and hole concentrations, respectively; V represents the local voltage; and D_n and D_p are the electron and hole diffusion coefficients, respectively.

$$\nabla^2 \phi = \frac{q}{\varepsilon_s}\left(N^+ - n + p\right) \quad (3)$$

In Eq. (3), ϕ is the potential; q is the electron charge; ε_s is the permittivity and N^+ denotes the free ionized carrier concentration.

$$\begin{aligned} I_{ds} &= C_{ox}\mu \frac{W}{L}\left(V_{gs} - V_{fb} - \Phi_F - \frac{V_{ds}}{2}\right)V_{ds} \\ &\quad - \frac{2}{3}\mu \frac{W}{L}\sqrt{2\varepsilon_s N^+ q} \\ &\quad \cdot \left(\left(2\Phi_F + V_{db}\right)^{3/2} - \left(2\Phi_F + V_{sb}\right)^{3/2}\right) \end{aligned} \quad (4)$$

Here, I_{ds} is the drain-source current; C_{ox} represents the capacitance of oxide layer; W and L are the gate width and length, respectively; V_{gs}, V_{ds}, V_{db} and V_{sb} are bias values between gate-source, drain-source, drain-back, source-back, respectively; V_{fb} is the flat band voltage; and Φ_F is the Fermi potential.

The device degradation simulation is based on the hot carrier induced oxide deterioration. Hot carrier injection is approximated using the "lucky electron" model [9, 10]. The location of peak injection probability in the channel is determined by the use of probability functions. Then, a deterioration model formulated by the use of short-term

stress experiments is applied to predict the change in local variables and device parameters. These changes can be represented using two different methods suggested in this work. The first method is based on imposing the shift in device performance on the properties of the oxide, thereby creating a virtual MOSFET whose behavior is equivalent to that of the initial device with degraded performance. The new oxide property is merely a time-average value of that property, which can be calculated using a simple average function (Eq. (5)). Since the change is induced by hot carrier injection, the local value of the shifted property will also be a function of the injection probability, P_{inj}.

$$\eta_{loc} = \frac{1}{t_w} \int_{t_0}^{t_0+t_w} \frac{\partial \eta\left(P_{inj}, E, T, \lambda, y\right)\Big|_{loc}}{\partial t} dt \qquad (5)$$

In Eq. (5), η_{loc} is the time average of the local oxide property to be calculated; P_{inj} the local hot carrier injection probability function derived from the lucky electron model; t_w the total work time; E the local electric field strength; T the local temperature; λ the mean free path; and y the distance into the oxide, normal to the oxide interface.

At this point, a second (virtual) MOSFET with the exact geometry is created numerically. The electrical and thermal properties of the new MOSFET's oxide layer are modified until the performance of the virtual device matches that of the degraded actual device (Eq. (6)). The electric field and current density are mainly a function of mobility, permittivity, and potential. One of these three is allowed to vary in order to reflect the changes in device behavior. Once the varying parameter is chosen and the change is calculated by the use of Eq. (5), the remaining two parameters are calculated to match the performance of the virtual device to that of the real one. In the expression below, the subscript *0* denotes the virtual device whose power output matches that of the degraded actual device, which is denoted by the subscript *deg*.

$$\begin{aligned} &E_0(\mu_0, \varepsilon_0, \phi_0) \cdot \vec{J}_0(\mu_0, \varepsilon_0, \phi_0) = \\ &\quad E_{\deg}(\mu_{\deg}, \varepsilon_{\deg}, \phi_{\deg}) \cdot \vec{J}_{\deg}(\mu_{\deg}, \varepsilon_{\deg}, \phi_{\deg}) \end{aligned} \qquad (6)$$

The second method is based on keeping the properties of the silicon dioxide layer fixed, but changing its thickness (h) or the effective gate length (L). Again, a second (virtual) MOSFET is created numerically. However, this time the properties of the virtual device are set equal to those of the actual device but the geometry (either h or L) is allowed to vary. After the varying parameter is chosen, the remaining one is calculated to match the performance of the virtual device to that of the real one. This is equivalent to creating a new oxide layer of a different thickness, which will cause the shift in behavior of the virtual MOSFET to match that of the degraded initial device (Eq. (7)).

$$\begin{aligned} &E_0(h_0, L_0) \cdot \vec{J}_0(h_0, L_0) = \\ &\quad E_{\deg}(h_{\deg}, L_{\deg}) \cdot \vec{J}_{\deg}(h_{\deg}, L_{\deg}) \end{aligned} \qquad (7)$$

Both methods introduced here can be used to simulate the degradation of a MOSFET. However, it is important to make sure that the device physics still hold while varying the parameters or geometry of the MOSFET. Neither the properties nor the geometry should be allowed to vary to the point where it will be necessary to use different governing equations than the ones here (too short or negative gate length, or impossibly high or low mobility or permittivity, etc.).

Using one of the two methods above, the coefficient ζ in the deterioration function is obtained. The device degradation in time can then be translated to a performance drop using a semi-empirical relationship (Eqs. (8a and 8b)). The deterioration coefficient ζ is affected by the ratio of lateral-vertical bias (V_{gb}/V_{ds}) and the gate oxide thickness (h). The gate length is orders of magnitude bigger than the oxide thickness. Therefore, the effect of the oxide thickness must be taken into account as a separate variable. The effect of gate length is negligible.

$$P_{\deg} = P_0 \exp(-\zeta\, t) \qquad (8a)$$

$$\zeta = \zeta\left(V_{gb}/V_{ds}, h\right) \qquad (8b)$$

The output power of the deteriorating device (P_{deg}) can thus be represented by an exponentially decaying function, where the degradation coefficient ζ is a complicated function of device parameters and properties; namely, the ratio of gate-back contact vs. drain-source bias and oxide thickness.

3. EXPERIMENTAL WORK

In order to establish a reliable database, a number of simple silicon-based MOSFETs were manufactured at the solid-state laboratory at SMU. Five 4-inch (p-type phosphorus-doped) silicon wafers were selected. Field and gate oxides were all thermally grown to 5 different thickness values (200-600 Å). All contacts were made of gold. To ensure good adhesion between the gold and oxide layers, a very thin layer (< 100 Å) of Cr was sputtered on the oxide before sputtering the gold. The thicknesses of sputtered and grown layers were all measured using a Dektak profiler to ensure that the thickness values are correct. The profiler has a measurement precision of ± 10 Å. The schematic of a

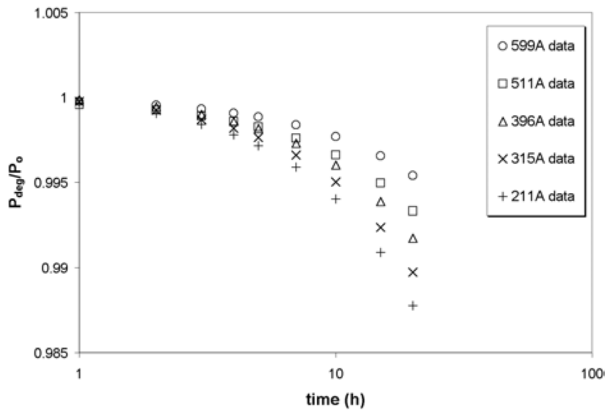

Fig. 2 Experimental results of short-term stressing induced by hot carrier injection in a MOSFET

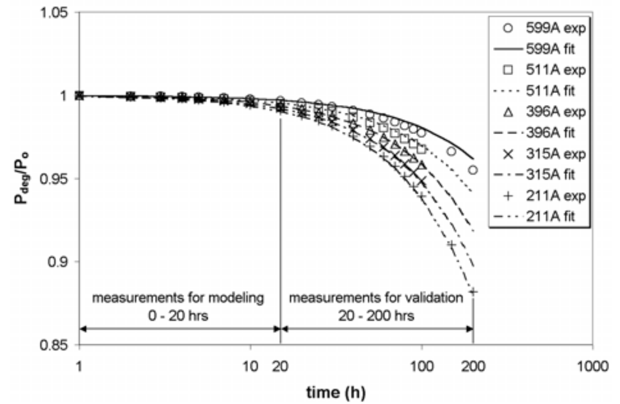

Fig. 3 Comparison of numerical and experimental degradation results

sample MOSFET used in this investigation is shown in Fig. 1.

The activation devices used in the electrical measurements consisted of: a high performance HP4142B power supply, with a resolution of 20 fA and 4 μV; an HP85120A drain pulser with a maximum duty cycle of 50%, a pulse repetition rate from single shot to 100 KHz and a pulse width from 250 ns to 1 ms; an HP6228A power supply capable of 50 W power output with up to 50 V or 2 A; and an HP8110A pulse generator for all standard waveforms with a maximum frequency of 150 MHz. The currents were delivered to the pads through 2.4 μm Tungsten needles. The combined uncertainty of the experiments was less than 5% for the power measurements.

On each wafer, devices of the same geometry (gate width, gate length, etc.) were selected for testing. Since all wafers of the same type were identical, the sole difference between two devices was the gate oxide thickness. Selected devices were then activated and stressed over two different periods of time. Short-term stressing was up to 20 hrs, whereas long-term was 20-100 hrs. In order to determine the drain-source bias to be used during stressing, identical MOSFETs were tested for their saturation and breakdown voltage. Since the MOSFETs used in the experiments are low power devices, a relatively low drain-source bias (80% of breakdown voltage) was chosen as the stressing bias. At certain times during stressing, the gate voltage was varied between the threshold voltage (typically 16-20 V) and a high level (typically 30-40 V). At each gate voltage, a set of output power measurements was taken. Then, an average and a standard deviation were calculated for each set. The highest standard deviation obtained for all the sets was ±2%.

4. RESULTS AND DISCUSSION

The experimental data obtained from the short-term (up to 20 hrs) stressing tests were used to determine the changes in MOSFET output power. That change was then mapped to a change in local variables and properties, which in turn was used to calculate the degradation coefficient (ζ) at each gate voltage. The degradation coefficients collected over the range of gate voltages were then used in the semi-empirical degradation module to obtain a general expression valid for all stressing periods. Results for selected gate voltages are listed in Table 1.

Table 1 Comparison of calculated and measured power output during degradation

h (± 25 Å)	V_{gb}/V_{ds}	ζ (1/hrs)	P_{deg}/P_o after 100 hrs (%, calculated)	P_{deg}/P_o after 100 hrs (%, measured)	Relative Difference (%)
211	4	0.00061	94.0	93.1	1.0
	6	0.00072	92.9	93.0	0.1
315	4	0.00050	95.0	94.5	0.6
	6	0.00060	94.1	94.0	0.1
396	4	0.00042	95.8	95.5	0.3
	6	0.00049	95.1	95.2	0.1
511	4	0.00031	96.9	97.1	0.2
	6	0.00036	96.3	97.0	0.7
599	4	0.00020	98.0	98.0	0.0
	6	0.00024	97.5	97.0	0.5

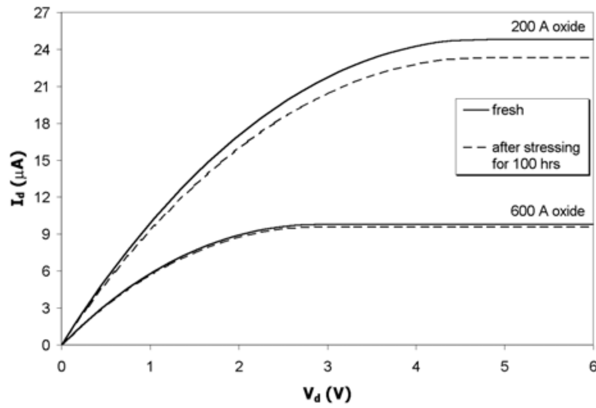

Fig. 4 Change in response of a degrading MOSFET with a 200 Å and 600 Å oxide layers

The degradation module modeled using the values obtained from evaluating the above short-term stressing data was then checked against the actual long-term (20-100 hrs) performance measurements, obtained through further stressing of the partially degraded MOSFET devices. The results from the long-term stressing experiments and the numerical degradation simulation were then compared to validate the model. In order to have a higher confidence level, two MOSFETs (one with a 200 Å and one with a 600 Å oxide layer thickness) were stressed for another 100 hrs and measured.

The experimental and numerical results were found to be in very good agreement for the 0-200 hr range for all test devices. Indeed, the degradation model based on coefficients calculated from the short-term stressing data predicts well the long-term degradation of the MOSFET (see Table 1). These results show that the degradation curve plotted using the model can be projected on a very long-term stress of MOSFETs to predict the useful lifetime of active devices. A comparison of the experimental and numerical results is plotted in Fig. 3.

Using the performance drop calculation, it is possible to determine when an active device within a circuit will have its power characteristics drop below a certain level. For the case when the circuit is digital, this information can then be used to predict its lifetime by setting a certain degradation limit at which the circuit is to be considered non-functional. For analog circuits, it is also useful to know what output power can be achieved by a stressed device at any given time. The I-V curves of a fresh and a stressed device are shown in Fig. 4. The responses shown are measured on two n-channel MOSFETs (one with 200 Å and the other with 600 Å thick gate oxides), activated with a gate bias of 20 V.

The degradation of the thinner oxide layer is more dramatic. This result means that devices with higher power output and thin oxide layer are more prone to oxide layer degradation and earlier breakdown if operated under high stress conditions.

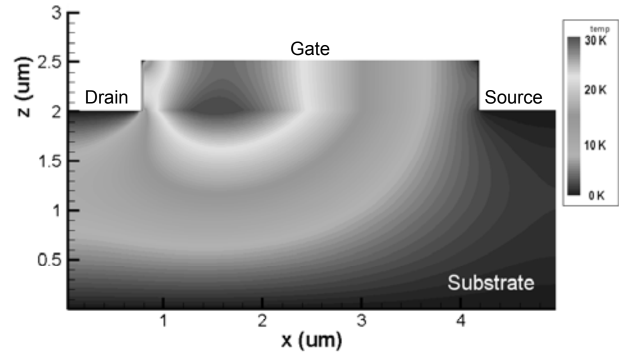

Fig. 5 Simulated temperature distribution on a fresh MOSFET

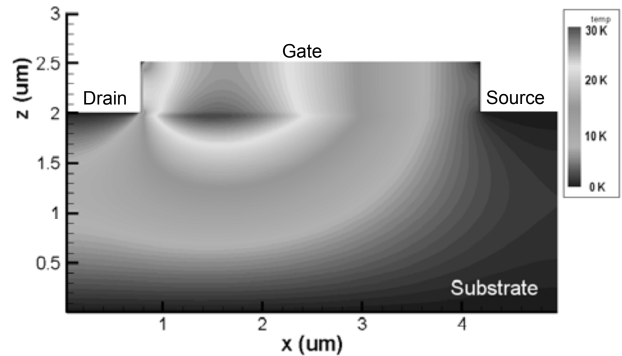

Fig. 6 Simulated temperature distribution on a degraded MOSFET (after 100 hrs)

The simulated temperature distributions for a fresh MOSFET and for a MOSFET that has been under hot carrier stress for 100 hrs are shown in Figs. 5 and 6, respectively. The stressed device has a lower peak temperature, which translates into lower power output under the same operating conditions. Because of the nature of the degradation function obtained in the previous section, the performance difference between the devices will grow exponentially in time. If there is a set condition for a performance drop in which the device can be deemed useful, the lifetime can be approximated as well. For instance if a designer accepts a 20% output power drop before deeming the device "dead", he can use this simulation to solve the reverse problem and determine the time when the power output drops below 80% of that of a fresh device.

5. CONCLUSIONS

This work introduced a new electro-thermal degradation simulation method and suggested its use in assessing the approximate lifetime of an active MOSFET under hot carrier injection stress. Experimental data from short-term stressing measurements were used to determine a degradation coefficient, which in turn was used as a basis to establish a model that is capable of predicting long-

term degradation effects and approximate lifetime of the device under test. The resulting model was checked against data from experiments of actual long-term MOSFET degradation and found to be accurate for the range of oxide thickness values considered in this investigation. In order to obtain a more complete theory, other device properties, such as doping concentration should be added. The outcome of this work is intended to contribute to the development of a simulation engine for coupled electro-thermal performance, which can be used to guide the analysis and design of high power density integrated circuits.

6. REFERENCES

[1] Diorio, C., Hasler, P., Minch, B., and Mead, C., "A complementary pair of four-terminal silicon synapses", *Analog Integrated Circuits and Signal Processing,* Vol. 13, pp. 153-166, 1997

[2] Cho, B., Xu, Z., Guan, H., and Li, M., "Effect of substrate hot-carrier injection on quasi-breakdown of ultrathin gate oxide" *Journal of Applied Physics,* Vol. 86, pp. 6590-6592, 1999

[3] Quader, K., Li, C., Tu, R., Rosenbaum, E., Ko, P., and Hu, C., "A new approach for simulation of circuit degradation due to hot-electron damage in NMOSFETs" *Int. Electron Devices Mtg. Tech. Digest*, p. 337, 1991

[4] Kuo, M., Seki, K., Lee, P., Choi, J., Ko, P., and Hu, C., "Simulation of MOSFET lifetime under AC hot electron stress" *IEEE Transactions on electron devices* Vol. 35, pp. 1004-1011, 1988

[5] Chung, J., Ko, P., and Hu, C., "A model for hot-electron induced MOSFET linear current degradation based on mobility reduction due to interface state generation" *IEEE Transactions on electron devices,* Vol. 38, pp. 1362-1370, 1991

[6] Glassbrenner, C. J., and Slack, G. A., "Thermal Conductivity of silicon and germanium from 3 K to the Melting Point," *Phys. Rev.*, Vol. 134, pp. A1058-A1069, 1964

[7] Streetman, B., *Solid State Electronic Devices,* 3rd ed., Prentice-Hall, New Jersey, 1990

[8] Zeghbroeck, B., *Principles of Semiconductor Devices*, web-book (http://ece-www.colorado.edu/~bart/book/), University of Colorado at Boulder, 1996

[9] Tam, S., Ko, P., and Hu, C., "Lucky electron model of channel hot electron injection in MOSFETs" *IEEE Transactions on electron devices,* Vol. 31, pp. 1116-1125, 1984

[10] Hu, C., "Lucky-Electron Model of Channel Hot Electron Emission," *Proceedings of Int. Electron Devices Meeting,* pp. 22-25, 1979