VLSI DESIGN
2001, Vol. 13, Nos. 1–4, pp. 111–115
Reprints available directly from the publisher
Photocopying permitted by license only

Impact of Scaling on CMOS Chip Failure Rate, and Design Rules for Hot Carrier Reliability

AMR HAGGAG[a,*], WILLIAM McMAHON[a], KARL HESS[a], BJÖRN FISCHER[a] and LEONARD F. REGISTER[b]

[a]*Beckman Institute, Department of Electrical and Computer Engineering and Department of Physics, University of Illinois, Urbana, IL 61801, USA;* [b]*Department of Electrical Engineering, University of Texas, Austin, TX 78712-1100, USA*

Silicon-hydrogen bonds passivate the interface defects at the silicon–silicon dioxide interface of CMOS transistors. The activation of these bonds and subsequent creation of interface traps is an important source of transistor degradation at current operating conditions. There is now evidence for a distribution in the activation energies of these bonds instead of a single threshold value. We show that conventional CMOS scaling rules are substantially affected by this energy distribution, as it causes an increased probability of smaller devices having lower activation thresholds and therefore faster activation times. Further, we quantify the voltage shift necessary to overcome the decreased yield due to the increased number of early device failures, and show, for 0.1 μm MOSFET scaling, that this shift can be a considerable fraction of the conventionally designed supply voltage.

Keywords: CMOS chip lifetime; Scaling and hot carrier reliability; Design rules for reliability; Sublinear lifetime dependence; Activation energy distribution

INTRODUCTION

There are at least two degradation mechanisms in MOSFETs that are caused by the hot carrier-induced activation of hydrogen from the silicon–silicon dioxide interface: the creation of interface traps, and the creation of bulk oxide traps [1]. The creation of interface traps is significant for MOSFET channel hot electron degradation [2, 3]. Interface traps are responsible for a reduction in the saturation drain current (I_{DSAT}) of the MOSFET which causes an increase in the propagation delay [4] which is then responsible for failure of the chip. The activation energies of the hydrogen bonds are therefore an important parameter for the degradation process. Note, that this may also be true for the creation of bulk oxide traps, not discussed here, which likewise seem to be related to hot carriers and hydrogen release [1].

Past models [2] have assumed that the hydrogen activation energy was constant at around $E_{A,IT}^{(mean)} \approx 3.7\,\mathrm{eV}$. This is inconsistent with

*Corresponding author. Fax: (217) 244-4333, e-mail: haggag@uiuc.edu

charge-pumping measurements, which show broadened energy distributions of the hot carrier induced defect states with half-widths σ larger than 0.1 eV [5]. Disorder-induced variations among the Si—H activation energies, $E_{A,IT}$, must be at least of this order, and are believed to be a source of the sub-linear time dependence of nMOSFET aging [5]. If we assume that the bond energies of individual bonds independently follow such a distribution, then each device will have a different set of activation energies, and we can expect variations in the hot carrier aging rate of individual nMOSFETs. These variations will result in a distribution in transistor and chip lifetimes. As the geometry of devices diminishes, there is an increased probability of having a device, or a critical path of devices, that has a sufficient number of hydrogen bonds with low activation energies to fail at much smaller times than the mean time to failure. In other words, the distribution of device lifetimes increases its width as we shrink the MOSFETs affecting device reliability. Quantifying this effect on the lifetime of the MOSFET is the subject of this paper.

ANALYTICAL CONSIDERATIONS

Measurements for I_{DSAT} degradation (our performance measure) under the worst-case stress show a corresponding lifetime t following an Arrhenius-type relation or models such as Ref. [2]:

$$\log t = \frac{ME_{A,IT}}{qV_{DD}} + C \tag{1}$$

where t is the time needed to activate Si—H bonds with activation energy less than $E_{A,IT}$ (enough for *specific* device failure). Charge-pumping measurements show a broadened fermi-derivative distribution[1] in the energies of the defect states resulting from the activation of the Si—H bonds at the interface. Similar disorder-induced variations in the Si—H bond energies can be expected (physical explanation analogous to theory in [6]):

$$f(E_{A,IT}) = \frac{1}{\sigma_{IT}} \frac{\exp\left(\left(E_{A,IT}^{(mean)} - E_{A,IT}\right)/\sigma_{IT}\right)}{\left[1 + \exp\left(\left(E_{A,IT}^{(mean)} - E_{A,IT}\right)/\sigma_{IT}\right)\right]^2} \tag{2}$$

The integral of this distribution is the probability that the fraction of activated Si—H bonds under hot-carrier stress have activation energy less than or equal to $E_{A,IT}$:

$$\mathcal{F}_{HCI} = \text{degradation} = \frac{\Delta D_{it}}{D_{it0}} \approx \frac{1}{1 + \exp\left(\left(E_{A,IT}^{(mean)} - E_{A,IT}\right)/\sigma_{IT}\right)} \tag{3}$$

where D_{it0} and ΔD_{it} are the original interface defect density and the hot carrier activated defect density, respectively. Then $(\Delta I_{DSAT}/I_{DSAT0})$ is a function of $\mathcal{F}_{HCI}$. Defining a lifetime constant τ_{IT} using:

$$\log \tau_{IT} = \frac{ME_{A,IT}^{(mean)}}{qV_{DD}} + C \tag{4}$$

Then we get by inserting Eqs. (1) and (4) into Eq. (3):

$$\mathcal{F}_{HCI} \approx \frac{1}{1 + (t/\tau_{IT})^{-\alpha_{HCI}}} \tag{5}$$

with:

$$\alpha_{HCI} \approx \frac{qV_{DD}}{\sigma_{IT} M \ln(10)} \tag{6}$$

The technology-dependent constant M can be extracted from the plot of $\log t_{measured}$ *versus* $(1/V_{DD})$ for a *measured* device (*e.g.*, we use data

[1] This form of the activation energy distribution has been used in the study of the reliability of grating-based optical dense-wavelength division multiplexed (DWDM) chip interconnects where bus signals are carried by separate lightwaves in the same link. The grating degradation mechanism (which affects optical signal to noise ratio, SNR_{opt}, in the links) involves decay of trapping centers (such as H_2 centers) in silica compounds [6, 7].

from [8]):

$$\log t_{\text{measured}} = \frac{ME_{\text{A,IT}}^{(\text{measured})}}{qV_{\text{DD}}} + C \qquad (7)$$

where t_{measured} is the time needed to activate Si—H bonds with activation energy less than $E_{\text{A,IT}}^{(\text{measured})}$ (enough for *measured* device failure). Assuming a starting number of N_H neutral defects (hydrogen) for this device, then the actual number of defects N_{f} needed for device failure (*i.e.*, 5% I_{DSAT} degradation) can be obtained from charge-pumping or simulation data and can be used to solve for $E_{\text{A,IT}}^{(\text{measured})} = \{E_{\text{A,IT}} | \mathcal{F}_{\text{HCI}}(E_{\text{A,IT}}) = N_f/N_H\}$ by using Eq. (3). In this work,[2] ρ_H is assumed to be $\approx 10^{12}\,\text{cm}^2$ in the whole channel, and we calculate the average number of silicon-hydrogen bonds which are in danger of activation by assuming a box of width W (width of the device) and length L_{it} at the drain end of the device. We define L_{it} as the region where the electric field drops to one fourth of its maximum value. The number of silicon-hydrogen bonds in that region is then $N_H = \rho_{\text{H}} L_{\text{it}} W$. Then we can calculate the number of bonds required to cause the device to fail, N_f, by performing $\text{DESSIS}_{\text{ISE}}$ TCAD simulation and find the density of interface traps ρ_{it} required to cause failure where failure is defined as a 5% degradation in I_{DSAT}. Throughout, we use the International Roadmap for Semiconductors and the well tempered MOSFET part of the DesCArtES project [10] to obtain values of various device parameters. Equation (6) gives a very important relationship between the sub-linear power-law factor α_{HCI} and V_{DD}/M, once σ_{IT}, the half-width of the defect activation energy distribution, is known. The above analytical form for the power-law factor (*i.e.*, Eq. (6)) can serve as a great tool for chip reliability engineers as discussed in [5] for the design and understanding of hot-carrier immune devices. One can replicate the measured sub-linear time-dependence of hot carrier aging with an appropriate choice of the half-width σ_{IT}. We show in [11], using a 180 nm device that $\sigma_{\text{IT}} \approx 0.2\,\text{eV}$. In addition, we have $E_{\text{A,IT}}^{(\text{mean})} \approx 3.1\,\text{eV}$.

We would like to add that we have described a model of the time-dependence of hot-carrier degradation *owing* to a unimodal distribution of hydrogen activation energies that gives the single power-law of Eq. (5). However, recent theoretical calculations on the $\text{Si}-\text{SiO}_2$ interface structure [11] suggest that defect activation can occur with probability p around a principal mean energy $E_{\text{A,IT}}^{(\text{mean1})} \approx 3.5\,\text{eV}$ as well as a lower secondary mean energy $E_{\text{A,IT}}^{(\text{mean2})} \approx 2.9\,\text{eV}$ with probability $1-p$ suggesting that the activation energy distribution is actually bimodal. As shown in [5], this would give a slightly more accurate time-dependence then Eq. (5) which is a double-power law:

$$\mathcal{F}_{\text{HCI}}(t) \approx (p/(1 + (t/\tau_1)^{-\alpha_1})) + ((1-p)/(1 + (t/\tau_2)^{-\alpha_2})) \qquad (8)$$

FAILURE STATISTICS AND RELIABILITY ASSURANCE

The failure probability of a device before time t (or equivalently by the activation of a **sufficient number** of hydrogen with activation energy less than or equal to $E_{\text{A,IT}}$, where t and $E_{\text{A,IT}}$ are related by the Arrhenius equation) is given by the finite binomial probability (after some mathematical manipulations):

$$F_{\text{HCI}}(E_{\text{A,IT}}(t)) = \sum_{n=N_f}^{n=N_H} \binom{N_H}{n} [\mathcal{F}_{\text{HCI}}(E_{\text{A,IT}})]^n [1 - \mathcal{F}_{\text{HCI}}(E_{\text{A,IT}})]^{N_H - n} \approx \sum_{n=N_f}^{n=N_H} \frac{[N_H \mathcal{F}_{\text{HCI}}(E_{\text{A,IT}})]^n \exp[-N_H \mathcal{F}_{\text{HCI}}(E_{\text{A,IT}})]}{n!} \qquad (9)$$

where $N_H = W\int_0^{L_{\text{eff}}} \rho_H(x)dx$ and $N_f = W\int_0^{L_{\text{eff}}} \rho_{\text{it}}(x)dx$ are the starting number of Si—H bonds at the interface and the number needed for 5% I_{DSAT} degradation respectively, $\rho_H(x)$ and $\rho_{it}(x)$ are the

[2] Work is under way to alternatively calculate $\rho_{it}(x)$ by using our newly developed interface state generation model with input of a nMOSFET hot carrier energy distribution given by the Monte Carlo code (MoCa) [9].

density of Si—H bonds and interface traps till failure, respectively, as a function of position x in the channel. We can extract the defect activation energy distribution from the model of Eq. (5) or Eq. (8) by fitting to short-time accelerated tests data [5]. Once we have determined the parameters of the activation energy distribution, we can assure reliability for all times by imposing the following constraints (with $MTF \geq (1/\alpha FIT)$):

$$\lambda(MTF) = (-(d \ln R(t))/dt)|_{MTF} \leq \alpha FIT \quad (10)$$

$$-\ln R(MTF) = \int_0^{MTF} \lambda(t)dt \leq (1/\beta)PPM \quad (11)$$

where PPM is the failed parts per million devices goal, FIT is the failure in billion hour time units goal, α is the relative failure rate per device normalized to the 180 nm technology and β is the relative increase in transistor number (inverse of critical dimension) normalized to the 180 nm technology. In this way, during device lifetime, $t_f \ll MTF$, the probability of even weakest device failing is miniscule. We are imposing a constraint on the device reliability function and its slope, both of which are important for the long-term reliability qualification of not only the hot-carrier failure mechanism but also other failure mechanisms. With the design aims of Eqs. (10) and (11), we can calculate maximum supply voltage reductions for safety. With the shrinking geometry of MOSFETs, there is an increased probability, F_{HCI}, of devices having enough low activation energy defects which necessitates more than the conventional 10% $V_{\mathrm{DD,max}}$ reduction for safety. If we require failure rates below 10 FIT as well as reliability below 50 PPM as design aim for all generations (Fig. 1), then much larger $\Delta V_{\mathrm{DD,max}}$ are required for smaller devices (Fig. 2).

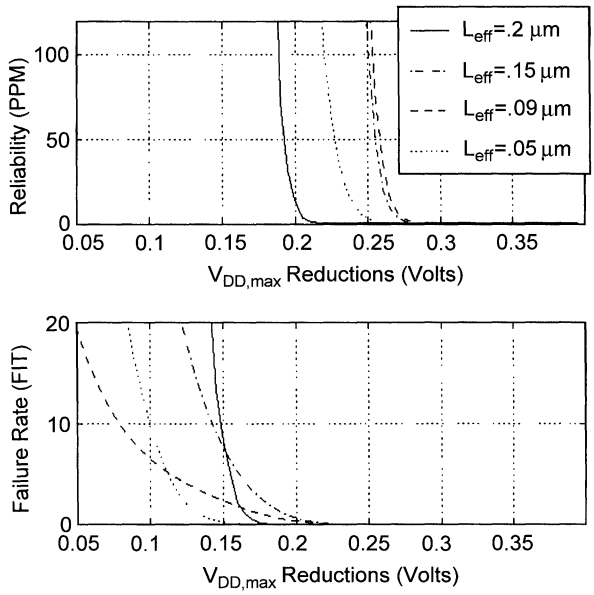

FIGURE 1 MOSFET FIT (inverse of MTF) and PPM (failed parts per million) *versus* maximum supply voltage ($V_{DD,max} = 1.1\ V_{DD,roadmap}$) reductions. Device is safe when it meets *both* PPM and FIT criteria. The different CMOS generations are characterized by their effective channel length and the corresponding poly gate lengths: (a) $LG = .25\,\mu$m (b) $LG = .18\,\mu$m (c) $LG = .13\,\mu$m (d) $LG = .09\,\mu$m.

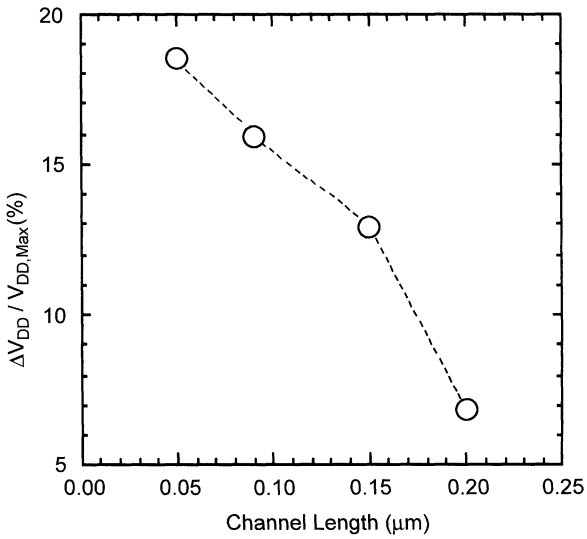

FIGURE 2 Reductions in maximum supply voltage, $V_{DD,max}$, for safety for the different technologies.

CONCLUSION

We have quantified CMOS chip failure rates from a first principles study of hot carrier aging of

nMOSFETs and by considering variations in such an aging rate due to a distribution in the width of the bond energies of passivating hydrogen at the interface. We emphasize that there are only three essential requirements for the effect discussed above to be significant in small geometry devices: (i) There must be a distribution of activation energies, although the specific form of the distribution is not critical, (ii) the number of charged defects necessary for degradation must decrease roughly as fast or faster than the channel width, and (iii) the fast degradation of a sizeable fraction of circuits is unacceptable. Each of these assumptions appears reasonable. Because of the effect of the bond energy distribution on the failure rate, we argue that the reliable operation of deep-submicron CMOS technology can only be achieved for $V_{DD,max} - \Delta V_{DD,max}$, where $V_{DD,max}$ can be determined from conventional methods of device reliability [8], and $\Delta V_{DD,max}$ can be determined from our results (Fig. 2). Another alternative would be the use of a better quality interface (such as deuterium passivated interfaces [3]) where lifetime variations are more tolerable.

Acknowledgement

We gratefully acknowledge Dr. Elyse Rosenbaum for valuable discussions. The work was supported by the Office of Naval Research and the Army Research Office.

References

[1] DiMaria, D. J. (1999). "Defect generation under substrate-hot-electron injection into ultrathin silicon dioxide layers", *Journal of Applied Physics*, **86**, 2100–2109.
[2] Hu, C. *et al.* (1985). "Hot-electron-induced MOSFET degradation-model, monitor, and improvement", *IEEE Transactions on Electron Devices*, **32**, 375–385.
[3] Hess, K. *et al.* (1998). "Giant isotope effect in hot electron degradation of metal oxide silicon devices", *IEEE Transactions on Electron Devices*, **45**, 406–416.
[4] Chen, K. *et al.* (1998). "Performance and V_{DD} scaling in deep submicrometer CMOS", *IEEE Journal of Solid State Circuits*, **33**, 1586–1589.
[5] Haggag, A. *et al.*, "High-performance chip reliability from short-time-tests: statistical models for optical interconnect and HCI/TDDB/NBTI deep-submicron transistor failures", *IEEE International Reliavility Physics Symposium*, Orlando, FL, April 30–May 3, 2001.
[6] Erdogan, T. *et al.* (1994). "Decay of ultra-violet induced fiber bragg gratings", *Journal of Applied Physics*, pp. 73–80.
[7] Haggag, A. *et al.*, "Performance degradation comparison between copper and optical interconnects in multi-GHz chips", *IEEE Journal on Lightwave Technology*, 2001, submitted.
[8] Li, E. *et al.*, "Hot carrier induced degradation in deep submicron MOSFETs at 100°C", *IEEE International Reliability Physics Symposium*, April 2000 and Personal Correspondence.
[9] Haggag, A., McMahon, W., Fischer, B. and Hess, K. unpublished.
[10] [www http://www.ceg.uiuc.edu/descartes.html], "Distributed Center for Advanced Electronics Simulation (DesCArtES)", 2000.
[11] Tuttle, B. *et al.* (1999). "Structure, energetics, and vibrational properties of Si—H bond dissociation in silicon", *Physical Review B*, **39**, 12884–12889.