The origin of aging in Al–SiO2–Si tunnel diodes

P. Chattopadhyay and K. Das

The origin of aging in Al–SiO$_2$–Si tunnel diodes

P. Chattopadhyay

Department of Electronic Science, 92, University College of Science, A.P.C. Road, Calcutta 700 009, India

K. Das

Department of Physical Chemistry, Indian Association for the Cultivation of Science, Jadavpur, Calcutta 700 032, India

The origin of aging of Al–SiO$_2$–Si tunnel diodes is investigated. A theoretical model is proposed based on the assumption that aluminum undergoes a chemical reaction with the SiO$_2$ layer and produces silicon which introduces active centers for trapping electrons in the oxide layer. The model shows that the negative charge density that develops at the interface is responsible for the time dependence of the barrier height observed in Al–SiO$_2$–Si tunnel devices. © *1996 American Institute of Physics.* [S0021-8979(96)05317-0]

The study of aging in metal–oxide–semiconductor (MOS) tunnel diodes is extremely important to ascertain the applicability of such devices for long-term operation. In several experimental studies, pronounced aging of MOS and related devices has been observed. Turner and Rhoderick[1] have noticed considerable aging of Cu, Ag, and Al contacts on chemically etched silicon surfaces. Card[2] observed pronounced aging of Al–SiO$_2$–Si tunnel diodes. Two possibilities are usually attributed to the above phenomena, namely, the interface contamination and the chemical reaction of metal with the oxide layer.[3]

One of the immediate consequences of the aging is either an increase (devices on n-type Si) or a decrease (devices on p-type Si) of the barrier height with time followed by a nearly constant value after a much longer time. Although the above trend has been consistently observed in many works, the theoretical understanding of this subject seem to be inadequate. Sharma and Srivastava,[4] in an attempt to explain Card's experimental data, assumed a reduction in oxide thickness due to the formation of a complex of aluminum oxide and silicon resulting from a chemical reaction between SiO$_2$ and Al. Based on an empirical relation between the complex layer and the time, the above workers interpreted the barrier height data. In this work we propose a different mechanism which offers a simple and better explanation for the aging effect in silicon MOS devices.

We stress that the aging effect is a manifestation of a chemical reaction between the metal and the adjacent oxide layer determined by the reaction rate. In an Al–SiO$_2$–Si system the aluminum reacts with silicon dioxide yielding the following products:

$$4\text{Al} + 3\text{SiO}_2 = 2\text{Al}_2\text{O}_3 + 3\text{Si}.$$
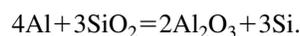
The above chemical reaction is consistent with the results of Auger electron analysis of the Al–SiO$_2$ interface which shows the presence of free Si, Al in Al$_2$O$_3$, and Si in SiO$_2$.[5] The possibility of such a reaction of metal with the interfacial layer has been corroborated by others.[3,4,6]

The reaction kinetics can be best understood in terms of a simple rate equation,

$$\frac{dN_{\text{SiO}_2}}{dt} = -r(t)N_{\text{SiO}_2}(t), \tag{1}$$

where N_{SiO_2} is the number of SiO$_2$ molecules present per unit area at any instant of time t for reaction with aluminum and $r(t)$ is the reaction rate. Note that the time dependence of the reaction rate will be determined by the growth kinetics of the Al$_2$O$_3$ layer which may be considered to be either linear or parabolic in time.[12] On integration of Eq. (1) one obtains

$$N_{\text{SiO}_2}(t) = N_{\text{SiO}_2}(0)\exp\left(-\int_0^t r(t)dt\right), \tag{2}$$

where $N_{\text{SiO}_2}(0)$ is the number of SiO$_2$ molecules per unit area available for reaction with Al at the Al–SiO$_2$ interface at time $t=0$. Therefore, the number of SiO$_2$ molecules consumed per unit area in time t is

$$N_{\text{SiO}_2}(0) - N_{\text{SiO}_2}(t)$$
$$= N_{\text{SiO}_2}(0)\left[1 - \exp\left(-\int_0^t r(t)dt\right)\right]. \tag{3}$$

The above quantity is equal to the number of silicon atoms liberated per unit area in time t [say, $N_{\text{Si}}(t)$]. One thus obtains

$$N_{\text{Si}}(t) = N_{\text{SiO}_2}\left[1 - \exp\left(-\int_0^t r(t)dt\right)\right]. \tag{4}$$

The functional form of the reaction rate $r(t)$ can be obtained by evaluating the flux of the Al atoms across the growing Al$_2$O$_3$ layer in manner similar to the one described by Grove[7] for atomic diffusion in solids in presence of an array of potential hills which impede the motion of the atoms. If there are n number of potential hills each of height W and width a, then the transfer probability of Al atoms to a SiO$_2$ site through the Al$_2$O$_3$ layer of thickness $X(t)$ is given by $\exp[-qWX(t)/akT]$. Thus, the number of atoms making such transitions per unit area is then $N_{\text{Al}}\exp[-qWX(t)/akT]$, where N_{Al} is the surface concentration of Al atoms at $x=0$. These atoms are then consumed by SiO$_2$ molecules at the rate $r(t)$. One therefore obtains

$$\frac{d}{dt}\left[N_{\mathrm{Al}}\exp\left(\frac{-qWX(t)}{akT}\right)\right]=-rN_{\mathrm{Al}}\exp\left(\frac{-qWX(t)}{akT}\right).$$

$$(5)$$

The above equation yields

$$r(t)=\alpha\,\frac{dX(t)}{dt},$$

$$(6)$$

where $\alpha=qW/akT$. On substituting Eq. (6) in Eq. (4), we obtain

$$N_{\mathrm{Si}}(t)=N_{\mathrm{SiO_2}}(0)[1-\exp\{-\alpha X(t)\}].$$

$$(7)$$

It therefore follows from the above equation that the growth of silicon surface concentration in the SiO$_2$ layer depends on the thickness of the growing Al$_2$O$_3$ layer.

The silicon atoms liberated at the Al–SiO$_2$ interface may have the properties of electron traps because of its tetravalent nature. Each of the atoms can accommodate a maximum of four electrons yielding a net negative charge density at the Al–SiO$_2$ interface. However, such incorporation of silicon atoms in the SiO$_2$ network may be accompanied by a partial saturation of dangling bonds and, thus, the released Si atoms cannot be treated as strictly fourfold electron traps. Also, the Si atoms may be lost due to the reverse reaction with Al$_2$O$_3$ layer. We accommodate the above two effects incorporating a lump parameter f, the maximum and minimum values of which are 4 and 0, respectively, and express the charge density trapped at the Si atoms as

$$Q_{\mathrm{Si}}=qfN_{\mathrm{SiO_2}}\{1-\exp[-\alpha X(t)]\}.$$

$$(8)$$

The charge neutrality condition in the presence of such silicon traps in an Al–SiO$_2$–Si system is given by

$$Q_m+Q_{\mathrm{it}}+Q_{\mathrm{sc}}+Q_f+Q_{\mathrm{Si}}=0,$$

$$(9)$$

where Q_{sc} is the depletion layer charge density, Q_{it} the interface-state charge density, Q_f the fixed oxide charge density, and Q_m the charge density that appears on the metal.

A correlation between the barrier height of the system (measured from the equilibrium Fermi level to the band edge at the interface) and various charge densities can be established following Cowley and Sze[8] and Refs. 9–11 given by

$$\phi_m-\chi-\phi_{bn}=\frac{\delta}{\epsilon_i}(Q_{\mathrm{sc}}+Q_{\mathrm{it}}+Q_f+Q_{\mathrm{Si}}),$$

$$(10)$$

where ϕ_m is the work function of the aluminum, χ the electron affinity of silicon, ϕ_{bn} the barrier height and δ is the SiO$_2$ layer thickness.

The interface-state charge density Q_{it} can be obtained following Cowley and Sze[8] given by

$$Q_{\mathrm{it}}=-q^2D_{\mathrm{it}}(E_g/q-\phi_0-\phi_{bn}),$$

$$(11)$$

where D_{it} is the interface state density and ϕ_0 is the pinning position of the Fermi level (relative to the valence-band edge at the interface) before the metal contact is established. With $Q_{\mathrm{sc}}=[2q\epsilon_sN_d(\phi_{bn}-V_n)]^{1/2}$ and using Eqs. (10) and (11), one obtains an expression for barrier height given by

$$\phi_{bn}=\phi+\left(\frac{C_2^2C_1}{2}-\frac{1}{2}(4\phi C_2^2C_1+C_2^4C_1^2-4C_2^2C_1V_n)^{1/2}\right),$$

$$(12)$$

where qV_n is the position of the Fermi level relative to the conduction band in the bulk,

$$\phi=\phi_{bn}(0)+\frac{4qfC_2\delta N_{\mathrm{SiO_2}}(0)}{\epsilon_i}\{1-\exp[-\alpha X(t)]\},$$

$$\phi_{bn}(0)=C_2(\phi_m-\chi)+(1-C_2)(E_g/q-\phi_0)+C_2\delta Q_f/\epsilon_i,$$

$$C_2=\epsilon_i/(\epsilon_i+q^2\delta D_{\mathrm{it}}),\quad\text{and}\quad C_1=2q\epsilon_sN_d\delta^2/\epsilon_i.$$

It can be shown that for low and moderate values of doping and ultrathin oxide layer, the contribution of the term within the parentheses is negligibly small and in such case the expression for barrier height reduces to

$$\phi_{bn}=\phi_{bn}(0)+C\{1-\exp[-\alpha X(t)]\},$$

$$(13)$$

where $C=4qfC_2\delta N_{\mathrm{SiO_2}}(0)/\epsilon_i$. In a similar way one can derive an expression for the barrier height for devices on p-type silicon given by

$$\phi_{bp}=\phi_{bp}(0)-C\{1-\exp[-\alpha X(t)]\},$$

$$(14)$$

where

$$\phi_{bp}(0)=C_2(E_g/q+\chi-\phi_m)+(1-C_2)\phi_0-C_2\delta Q_f/\epsilon_i.$$

The explicit time dependence of Eqs. (13) and (14) can be obtained if the functional form of $X(t)$ is known. We consider here well-known linear and parabolic law for oxidation of metals[12] and silicon[7] and replace $X(t)$ in Eq. (7) by A_lt or $A_p\sqrt{t}$, where A_l and A_p are, respectively, the linear and parabolic rate constants. Thus, the explicit time dependences of the barrier heights for n- and p-type devices for linear and parabolic growth law become

$$\phi_{bn,p}=\begin{cases}\phi_{bn,p}(0)+C[1-\exp(\beta_lt)]\\\phi_{bn,p}(0)+C[1-\exp(\beta_p\sqrt{t})],\end{cases}$$

$$(15)$$

where $\beta_l=\alpha A_l$ and $\beta_p=\alpha A_p$. Note that for the small values of $\beta_{l,p}$, the values of $\phi_{bn,p}$ reduce to $\phi_{bn,p}(0)$, respectively; however, for moderate and large values of $\beta_{l,p}$, the barrier height is sensitive to time. Also, from Eq. (6), the reaction rates are, respectively, $r_l=\beta_l$ and $r_p=\beta_p/2\sqrt{t}$ for the linear and parabolic growth law.

It is evident from Eq. (15) that the barrier height for n-type silicon ϕ_{bn} increases with time from an initial value $\phi_{bn}(0)$ until it becomes constant at $\phi_{bn}(0)+C$ after a much longer time. The nature of variation of barrier height for p-type silicon is just the opposite. According to Eq. (15), the barrier height decreases with time from a value $\phi_{bp}(0)$ and ultimately saturates at a value $\phi_{bp}(0)-C$. The variation of barrier height with time for Al–SiO$_2$–$n(p)$-Si diodes is shown in Fig. 1 for different values of linear and parabolic rate constants. The values of $\phi_{bn}(0)$, $\phi_{bp}(0)$, and C are considered to be 0.45, 0.75, and 0.25 eV, respectively. The experimental barrier height data[2] of 150 °C heat-treated Al–SiO$_2$–$n(p)$-Si diodes ($\delta=20$ Å) corrected for tunnel exponent are compared with the theory. It is seen that the
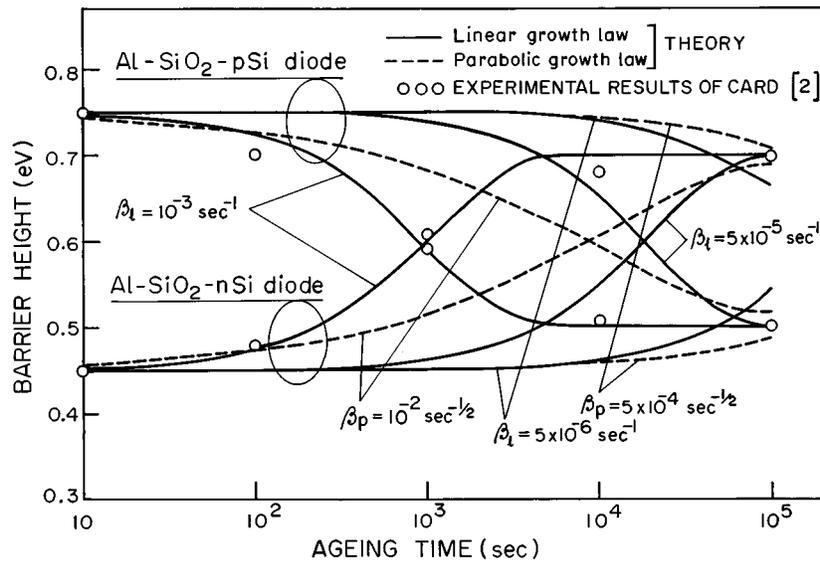
FIG. 1. Time dependence of barrier height of Al–SiO$_2$–n-Si and p-Si tunnel diodes. Parameters: $\phi_{bn}(0)=0.45$ eV; $\phi_{bp}(0)=0.75$ eV; and $C=0.25$ eV.

above experimental results can be best explained in terms of a linear rate constant $\beta_l=10^{-3}$ s^{-1} for both n- and p-type devices.

An estimation of the factor f can be made from the value of C with $C=0.25$ eV, $\delta=20$ Å, $N_{SiO_2}(0) = \rho_{SiO_2}\delta$, the density of SiO$_2$ molecules $\rho_{SiO_2} = 2.2 \times 10^{22}$ cm^{-3},[7] $D_{it}=2\times10^{12}$ cm^{-2} eV^{-1} for chemically prepared surface,[13] $\epsilon_i=8.854 \times10^{-14}$ F cm^{-1} (assumed to be the free space permittivity for ultrathin oxide layer), one obtains $f=5.35\times10^{-5}$. This implies that, due to the partial saturation of the the dangling bonds and the reversibility of the reaction, the maximum value of the charge density expected from the released Si atoms has been considerably reduced; however, even in that case the effective trapped charge density may cause considerable aging effect. Because of the above change in interface charge density with the growing Al$_2$O$_3$ layer, one cannot possibly rule out a flatband voltage shift in the $C-V$ characteristics of the aging devices.

The value of the rate constant A_l cannot be directly estimated from β_l since the latter constant involves the parameter α; however, an estimation of the rate constant is possible from the analysis of the weight gain of the aging devices. The method has been preposed by Gulbransen and Wysong,[14] to study the growth kinetics of the Al$_2$O$_3$ layer on Al substrate at elevated temperature. The observed growth of the Al$_2$O$_3$ layer was found to be either linear or parabolic in time depending on the growth conditions. The activation energy was estimated to be 22 800 cal/mol. We have consid-

ered here only those experimental data from the above reference for which the growth kinetics are linear and found the rate constant to vary in the range 2.1×10^{-1}–2.5×10^{-2} Å s^{-1} under various growth conditions. Considering the values of activation energy and the distance between the adjacent Al atoms (2.74 Å) reported in the above reference, we obtain $\alpha=9.91$ Å^{-1} and $A_l=1.01\times10^{-4}$ Å s^{-1} corresponding to the value of $\beta_l=10^{-3}$ s^{-1}. It therefore follows that the reaction of Al with the SiO$_2$ layer is much slower compared to normal oxidation of aluminum reported in Ref. 14. The reason for low value of A_l is possibly the temperature, which in the present case is only 150 °C, much less compared to the oxidation temperature reported in Ref. 14.

[1] M. J. Turner and E. H. Rhoderick, Solid State Electron. **11**, 291 (1968).
[2] H. C. Card, IEEE Trans. Electron Devices **ED-23**, 538 (1976).
[3] E. H. Rhoderick and R. H. Williams, *Metal–Semiconductor Contacts*, 2nd ed. (Clarendon, Oxford, 1988), p. 54.
[4] K. K. Sharma and G. P. Srivastava, Phys. Status Solidi A **89**, 403 (1985).
[5] Y. E. Strausser and K. S. Majumder, J. Vac. Sci. Technol. **15**, 238 (1978).
[6] T. W. Hickmott, J. Appl. Phys. **51**, 4269 (1980).
[7] A. S. Grove, *Physics and Technology of Semiconductor Devices* (Wiley, New York, 1967).
[8] A. M. Cowley and S. M. Sze, J. Appl. Phys. **36**, 3212 (1965).
[9] D. L. Pufrey, IEEE Trans. Electron Devices **ED-23**, 587 (1976).
[10] P. Chattopadhyay and A. N. Daw, Solid State Electron. **28**, 831 (1985).
[11] P. Chattopadhyay and K. Das, Solid State Electron. **34**, 1367 (1991).
[12] N. Cabrera and N. F. Mott, Rep. Prog. Phys. **12**, 163 (1948).
[13] H. C. Card and E. H. Rhoderick, J. Phys. D **4**, 1587 (1971).
[14] E. A. Gulbransen and W. S. Wysong, J. Phys. Colloid Chem. **51**, 1087 (1947).