Oxygen Vacancies in High Dielectric Constant Oxide-Semiconductor Films

Supratik Guha and Vijay Narayanan

IBM T. J. Watson Research Center, P.O. Box 218, Yorktown Heights, New York 10598, USA (Received 19 October 2006; published 8 May 2007)

We provide evidence that the oxygen vacancy is a dominant intrinsic electronic defect in nanometer scaled hafnium oxide dielectric films on silicon, relevant to microelectronics technology. We demonstrate this by douglaring a general model for the kinetics of owners usegney formation in match ultrathin oxide

this by developing a general model for the kinetics of oxygen vacancy formation in metal-ultrathin oxidesemiconductor heterostructures, calculating its effect upon the band bending and interfacial oxidation rates and showing good experimental agreement with the predictions.

DOI: 10.1103/PhysRevLett.98.196101

PACS numbers: 68.55.Ln, 71.55.-i, 77.55.+f, 85.30.Tv

Silicon-metal-oxide-semiconductor field effect transistor (MOSFET) technology today faces its biggest change with the impending replacement of the SiO₂ gate insulator with the higher dielectric constant (κ) ionic oxide HfO₂ [1]. A typical transistor "high- κ gate stack" consists of a $Si/\sim 1 \text{ nm } SiO_2/2-3 \text{ nm } HfO_2/metal electrode hetero$ structure where the quality of the HfO₂ and the thickness of the underlying SiO₂ are both critical. Progress, however, has been stymied by the recent observation of unexpected band bending in the silicon layer beneath the HfO_2 [2–4], manifested by threshold voltage (V_t) shifts of the MOSFET devices. These shifts indicate the presence of an unexplained positive charge or dipoles within the dielectric layer, which alters the electrostatic potential in the silicon layer adjacent to the oxide. The developments bear an interesting parallel to voltage shifts [5] that plagued SiO₂ MOSFETs 40 years ago, related in that case to the extrinsic contamination of SiO₂ with alkali ions. In contrast, the defects in HfO₂ are speculated to be *intrinsic* in nature, specifically positively charged oxygen vacancies [2,6,7]. This is based upon observed variations of the silicon band bending with varying oxygen impurities during anneals [2,3] and in concurrence with the defect chemistry literature of bulk ionic oxides such as HfO_2 and ZrO_2 [8,9]. However, alternate mechanisms have been suggested as well [10,11], and the matter has remained a speculation, since explicit experimental evidence for their presence and role specifically in nanometer scaled ionic metal-oxidesemiconductor MOSFETs has been lacking.

In this Letter, using a simple rate kinetic approach, we derive a relation for the expected concentration of the native oxygen vacancy defect in a silicon/thin metal oxide dielectric/metal electrode "gate stack," as a function of the oxygen partial pressure. We then determine how the vacancy concentrations will affect the threshold voltage shifts and the growth of the underlying interfacial silicon oxide layer. We experimentally measure these quantities as a function of oxygen partial pressure and find good agreement with the predictions of the model. These results conclusively establish the presence of charged oxygen vacancy defects as a significant native defect in this technologically important material.

Samples used were field oxide isolated 3 nm HfO₂/ 20 nm Re metal-oxide-semiconductor (MOS) capacitor structures grown on a 200 mm Si(3×10^{15} cm⁻³, *n*-type)/ ~1 nm SiO₂ wafer [Fig. 1(a)]. Samples were grown by chemical vapor deposition using a hafnium tertbutoxide precursor. Following an initial forming gas anneal at 500 °C for 30 minutes, oxygen anneals were carried out for 6 hours between 220 and 450 °C, at oxygen pressures



FIG. 1. (a) Schematic of the silicon-metal oxide-metal gate stack, showing the mechanism of transport of oxygen from the ambient to the silicon interface; (b) experimental capacitance vs voltage curves obtained across this stack after anneals (6 h at 400 °C) at different partial pressures of oxygen. Note the parallel shift along the voltage axis (indicated by the line with arrows) of ~ 0.33 V with increasing oxygen pressure. This shift is tracked by noting the voltage ϕ^* at a fixed capacitance of 5×10^{-3} F/cm².

varying from $<10^{-6}$ to 2×10^{-2} torr in a high vacuum chamber. Electronic properties of the MOS stacks were determined by capacitance voltage (CV) measurements at a frequency of 100 kHz. By sweeping the voltage applied to the metal electrode, the electrostatic potential in the silicon adjacent to the dielectric is altered such that this region goes from electron accumulation to electron depletion and the formation of a space charge region, yielding the characteristic CV curve. In a simple model, the position of the CV curve along the voltage axis is determined by the offset of the Fermi levels between the metal and the Si layer. The presence of a fixed charge in the oxide, however, creates a potential drop across it, resulting in a parallel translation of the CV curve along the voltage axis by an equal amount. This shift can thus be tracked by noting the voltage (ϕ^*) at a specific value of the capacitance in a CV curve and is identical to the threshold voltage shift [12] that would have been observed were a MOS transistor to be tested. The peak of the capacitance per unit area (C_a) , when the Si layer is in accumulation, is used to extract the equivalent electrical thickness of the dielectric stack $t_{\rm el} = \epsilon_r \epsilon_0 / C_a$, where ϵ_r is the dielectric constant of SiO₂ and ϵ_0 is the vacuum permittivity. The quantity $t_{\rm el}$ is the electrically determined average equivalent thickness of the stack. Relative differences in $t_{\rm el}$ of ~0.1 nm can be discriminated experimentally.

The CV curves for a set of wafers as a function of different oxygen partial pressures is shown in Fig. 1(b). First, for anneals at increasing partial pressures of oxygen, the curves are shifted to the right by as much as ~ 0.5 V, consistent with a decrease in the net positive charge in the dielectric. In order to quantify this shift, we have plotted the voltage ϕ^* measured at a fixed capacitance of 5 \times 10^{-3} F/cm² as a function of the oxygen partial pressure during anneals at different temperatures [Fig. 2(a)]. For Re electrodes (work function of ~5 eV), one would expect ϕ^* to be around 1 V if there was no fixed charge present in the dielectric. A second key feature is the drop in the accumulation capacitance as a function of anneals at increasing oxygen partial pressures-an effect well known to be caused by the growth of the interfacial silicon oxide layer [Fig. 1(a)] due to oxygen diffusion through the gate stack [13-17]. The interfacial growth is shown in Fig. 2(b), where we plot the electrical thickness (t_{el}) of the dielectric stack, as a function of the oxygen pressure. Returning to Fig. 1(b), at intermediate voltage biases between depletion and accumulation, the capacitance data typically exhibit a shoulder. This is invariant to frequency variations and is unrelated to the formation of interfacial traps. A detailed discussion of this is beyond the scope of this Letter [18], and it does not affect the two key observations for this Letter. The hysteresis observed due to voltage stressing in the double sweeps of Fig. 1(b) are ~ 10 mV or less.

Consider the process of oxygen incorporation for the heterostructure stack shown in Fig. 1(a). It is well known that dissolved oxygen in the metal electrode [14-16] acts as a source of oxygen for growth of the interfacial silicon



FIG. 2. (a) The translation of the capacitance voltage curves as a function of oxygen partial pressure (p_{O_2}) during anneals at different temperatures is shown by plotting the voltage ϕ^* versus p_{O_2} , where ϕ^* is measured at a fixed capacitance of 5×10^{-3} F/cm² from the different CV traces; (b) the variation of the electrical thickness (t_{el}) as a function of p_{O_2} during anneals at different temperatures. Anneal temperatures are 450 (solid circles), 400 (open circles), 260 (solid squares), and 220 °C (open squares). The drawn curves are for guiding the eyes only.

oxide layer and that the overlying metal oxide layer mediates atomic oxidation of the silicon [13,16]. We view the process of oxygen transfer and interfacial SiO₂ growth as a sequence of the following steps: (i) the incorporation of oxygen from the ambient $[O_2(g)]$ into the metal layer (O_M) , $\frac{1}{2}O_2(g) = O_M$; (ii) transfer of the oxygen atom from the metal to an oxygen site (O_a) in the metal oxide by annihilation of an oxygen vacancy (V_o) , $V_0^{-} + 2e +$ $O_M = O_O^X$ or $V_O^X + O_M = O_O^X$, depending upon whether the vacancy is charged or neutral; (iii) migration of the oxygen atom by vacancy exchange to the metal oxide/ SiO₂ interface; (iv) transfer of atomic oxygen from the metal oxide into the SiO₂ layer with the return of an oxygen vacancy to the metal oxide and oxidation of the underlying silicon, $\frac{1}{2}$ Si + O₀^X = V_0^X [or $(V_0^{..} + 2e) + \frac{1}{2}$ SiO₂]. Here the superscripts "x" and "··" refer to a neutral or doubly positively charged species, respectively. The fraction of oxygen vacancies that are positively charged can contribute to the voltage drop across the dielectric metal oxide layer, as will be described below.

If the metal is thin enough (< few tens of nanometers) such that it rapidly equilibrates with the ambient oxygen at the temperature of interest, we can write, following the reaction in step (i), for the concentration of dissolved oxygen in the metal (c_M) up to the terminal solubility [19]:

$$c_M = \frac{p_{O_2}^{1/2}}{\gamma}.$$
 (1)

Here c_M is expressed as a site fraction, relative to the number of available oxygen interstitial sites (N_M) , γ is the activity coefficient, and p_{O_2} is the oxygen partial pressure. If $[V_o]$ is the fraction of oxygen sites in the metal oxide that are vacant, then, assuming the layer is thin enough that the oxygen vacancy concentration within it is uniform, the annihilation rate of the vacancies is given by

$$\ell_{\text{HfO}_2} N_o \frac{d[V_o]}{dt} = -\alpha a N_M c_M [V_o] - Ba N_{\text{SiO}_2} [V_o] + A N_o a + \beta N_o a.$$
(2)

Here α , β , A, and B are temperature-dependent rate constants, N_o and N_{SiO_2} are the number of oxygen sites per unit volume in the HfO₂ and SiO₂, respectively, "*a*" is of the order of a lattice constant, and l_{HfO_2} is the HfO₂ layer thickness. The first and second terms on the right-hand side of Eq. (2) are the rates at which oxygen atoms migrate into the metal oxide from the metal electrode and from the SiO₂ layer, respectively. The third and fourth terms are the rates at which oxygen atoms migrate into the SiO₂ and the metal, respectively, from the HfO₂. We have made the approximation $AN_o(1 - V_o^{-})a \sim AN_oa$, and $\beta N_o(1 - V_o^{-})a \sim \beta N_oa$. Combining with Eq. (1), and making the approximation $N_o \sim N_{\text{SiO}_2} \sim N_M = N$, the solution to Eq. (2) is given by

$$[V_o] = \left([V_{o,\text{ini}}] - \frac{A + \beta}{\alpha \gamma^{-1} p_{O_2}^{1/2} + B} \right) e^{-\operatorname{ta}(\alpha \gamma^{-1} p_{O_2}^{1/2} + B)/\ell_{\text{HfO}_2}} + \frac{A + \beta}{(\alpha \gamma^{-1} p_{O_2}^{1/2} + B)},$$
(3)

where $V_{o,\text{ini}}$ is the initial value of the oxygen vacancy concentration, at t = 0.

Consider the electrical effects of oxygen vacancies. Neutral oxygen vacancies can behave as electron donors in oxides, via the process $V_o^x = V_o^{-} + 2e$, and theoretical estimates place their energy level at ~3.8 eV above the HfO₂ valence band [7]. Referring to the heterostructure in Fig. 1, if the metal electrode Fermi level is below the electronic level of the neutral O vacancy, the electrons can tunnel out of the dielectric and into the metal [2,6]. The dielectric layer then acquires a net positive charge and a resultant potential drop across itself. From a solution of the Poisson equation for uniformly distributed charged vacancies [20], this potential drop, which is equal to the lateral translation (in voltage) of the CV curves, is given by

$$\phi_{\rm id} - \phi^* = \Delta \phi = \frac{e N_{\rm O} \theta [V_o] \ell_{\rm HfO_2}^2}{\varepsilon_r \varepsilon_0}.$$
 (4)

Here *e* is the electron charge, θ is the fraction of oxygen vacancies that are charged, ϕ^* is the voltage at a fixed value of the capacitance from the CV curves, with $\phi^* = \phi_{id}$ for the ideal case, when there are no charge defects in the oxide. Substituting the value of $[V_o]$ from Eq. (3) into Eq. (4), one obtains the dependence of $\Delta\phi$ on the oxygen partial pressure. We simplify Eq. (3) further for two limiting conditions. When $ta_M(\alpha\gamma^{-1}p_{O_2}^{1/2} + B) \gg \ell_{HfO_2}$, we recover the steady state solution of Eq. (2), with $dV_o/dt = 0$. With the further approximation $\alpha\gamma^{-1}p_{O_2}^{1/2} \gg B$, valid when more oxygen is injected into the metal oxide from the metal electrode than from the SiO₂ layer, we can combine Eqs. (3) and (4) to obtain

$$\Delta \phi = \frac{K_1}{p_{\mathrm{O}_2}^{1/2}},\tag{5}$$

where

$$K_1 = \frac{eN_{\rm O}\theta\ell_{\rm HfO_2}^2(A+\beta)}{\varepsilon_r\varepsilon_0\alpha\gamma^{-1}}$$

This steady state solution will be valid at higher temperatures where the rate constant α is high, and the steady state condition is reached quicker. Under these conditions, expected at higher temperatures, a plot of $\ln(\Delta \phi)$ vs $\ln(p_{O_2})$ should exhibit a slope of ~ -0.5 .

Under transient conditions, when $A + \beta \ll \alpha \gamma^{-1} p_{O_2}^{1/2} + B$, we can simplify Eq. (3) and combine with (4) to obtain

$$\ln(\Delta\phi) = K_2 - \frac{\tan\alpha\gamma^{-1}p_{O_2}^{1/2}}{\ell_{HfO_2}},$$
 (6)

where

$$K_{2} = \ln\left(\frac{eN_{O}\theta\ell_{HfO_{2}}^{2}V_{o,\text{ini}}}{\varepsilon_{r}\varepsilon_{0}}\right) - \frac{\mathrm{ta}B}{\ell_{HfO_{2}}}$$

is a constant. This will be valid at low temperatures when the rate constants A and β are small and the partial pressure of oxygen is high enough. In this case, a plot of $\ln(\Delta \phi)$ vs $(p_{O_2})^{0.5}$ will be linear, with a slope of $-(t\alpha \gamma^{-1}a/\ell_{\rm HO_2})$.

Let us now turn to the interfacial silicon oxide growth rate dependence. At the very small (~ nanometer) thicknesses of the interfacial SiO₂ and the HfO₂, it is reasonable to expect that the growth rate $(d\ell_{SiO_2}/dt)$ of the SiO₂ layer is not limited by oxygen transport through the layer thickness but by the reaction kinetics controlling the net flux of oxygen atoms exchanged across the SiO₂/HfO₂ interface. Then

$$\frac{d\ell_{\rm SiO_2}}{dt} = \frac{\Omega}{2} (AN_o a - BN_{\rm SiO_2}V_o a) \sim \frac{\Omega N a}{2} (A - BV_o).$$
(7)

Under steady state conditions $(dV_o/dt = 0)$, we can combine the solution for Eq. (7) with Eq. (4) to obtain the relation between the interfacial oxide thickness ℓ_{SiO_2} and



FIG. 3. The translation in the capacitance voltage curves $(\Delta \phi)$, plotted in (a) as a function of the oxygen partial pressure (p_{O_2}) for the samples annealed at 450 (triangles) and 400 °C (squares), (b) versus $(p_{O_2})^{0.5}$ for the samples annealed at 260 (circles) and 220 °C (open squares), and (c) versus the electrical thickness of the stack (t_{el}) for samples annealed at 400 (solid circles) and 450 °C (open circles).

the voltage shift $\Delta \phi$:

$$\Delta \phi = -\frac{2e\ell_{\rm HfO_2}^2}{at\Omega B\varepsilon_r \varepsilon_0} \ell_{\rm SiO_2} + C, \qquad (8)$$

where *C* is a constant. A plot of ℓ_{SiO_2} vs ϕ^* will therefore exhibit a linear relationship under these conditions. This will be likely to be true at higher temperatures.

We now turn to a comparison of the above three predictions to the data. The variation of ϕ^* with p_{O_2} at 400 and 450 °C compared to the data at 220 and 260 °C in Fig. 2(a) indicates different functional dependences and, as we show below, is in good agreement with the models described above for the steady state and transient cases. In Fig. 3(a), we have replotted the data at 400 and 450 °C, in terms of $\ln(\Delta \phi)$ vs $\ln(p_{O_2})$. We take ϕ_{id} to be ~1.1 V, i.e., the asymptotic value for ϕ^* at high values of p_{O_2} [from Fig. 2(a)]. The data exhibit linear fits to slopes of ~ 0.38 and 0.42, respectively, in good agreement with the predicted slope of 0.5 from Eq. (5) for the steady state case, expected at higher temperatures and reflecting the $(p_{O_2})^{0.5}$ dependence. The small discrepancy is possibly related to the assumptions made in arriving at Eq. (5). Clearly, the nanometer scale thicknesses of the dielectric and metal layers allow a steady state to be established at temperatures as low as ~ 400 °C. The lower temperature data (220 and 260 °C anneals), shown in Fig. 3(b), are plotted as $\ln(\Delta \phi)$ vs $(p_{O_2})^{0.5}$ and are in good agreement with the linear dependence expected from Eq. (6) for the transient case. Finally, Eq. (8), based upon the steady state assumption, predicts that the plot of ϕ^* vs interfacial oxide thickness should be linear. The experimental data for the high temperature samples, plotted in Fig. 3(c), exhibit excellent linearity in agreement with this prediction. This is a key agreement, since the only assumption here is one of steady state.

The above model based upon oxygen vacancies and their experimental validation therefore uniquely ties together the apparently disparate observations of interfacial oxide growth and band bending changes as a function of annealing in oxygen ambients. The results provide conclusive evidence that charged intrinsic oxygen vacancies are a significant defect in the HfO_2/Si system.

One of us (S.G.) thanks Dr. E. Tutuc and Dr. M. Ghosh for discussions.

- S. B. Samavedam *et al.*, Int. Electron Devices Meet. Tech. Dig. **2002**, 433 (2002).
- [2] E. Cartier et al., in 2005 Symposium on VLSI Technology. Digest of Technical Papers (IEEE, New York, 2005), p. 230.
- [3] J.K. Schaeffer et al., Appl. Phys. Lett. 85, 1826 (2004).
- [4] D. Lim et al., Appl. Phys. Lett. 87, 072902 (2005).
- [5] E.H. Snow et al., J. Appl. Phys. 36, 1664 (1965).
- [6] K. Shiraishi et al., Jpn. J. Appl. Phys. 43, L1413 (2004).
- [7] K. Xiong et al., Appl. Phys. Lett. 87, 183505 (2005).
- [8] P. Kofstad and D.J. Ruzicka, J. Electrochem. Soc. 110, 181 (1963).
- [9] See, for example, W.E. Wang and D.R. Olander, J. Am. Ceram. Soc. **76**, 1242 (1993), and references therein.
- [10] Y.C. Yeo et al., J. Appl. Phys. 92, 7266 (2002).
- [11] L. Pantisano, Appl. Phys. Lett. 88, 243514 (2006).
- [12] The threshold voltage is the gate bias at the onset of carrier inversion in a MOS transistor.
- [13] V. Narayanan et al., Appl. Phys. Lett. 81, 4183 (2002).
- [14] R. P. Pezzi et al., Appl. Phys. Lett. 88, 243509 (2006).
- [15] E. J. Preisler et al., Appl. Phys. Lett. 85, 6230 (2004).
- [16] B. W. Busch et al., Phys. Rev. B 62, R13 290 (2000).
- [17] V.K. Paruchuri, Appl. Phys. Lett. (to be published).
- [18] A.C. Callegari (unpublished).
- [19] Y.A. Chang et al., Prog. Mater. Sci. 32, 97 (1988).
- [20] R.S. Muller and T.I. Kamins, *Device Electronics for Integrated Circuits* (Wiley, New York, 1986), p. 399, 2nd ed.