Origin of flat-band voltage sharp roll-off in metal gate/high-k/ultrathin-SiO2/Si p-channel metal-oxide-semiconductor stacks

X. H. Zheng,1 A. P. Huang,1,∗ Z. S. Xiao,1 Z. C. Yang,1 M. Wang,1 X. W. Zhang,2 W. W. Wang,3 and Paul K. Chu4
1Department of Physics, Beihang University, Beijing 100191, People’s Republic of China
2Key Laboratory of Semiconductor Materials Science, Institute of Semiconductors, Chinese Academy of Sciences, P.O. Box 912, Beijing 100083, People’s Republic of China
3Integrated Circuit Advanced Process Center, Institute of Microelectronics, Chinese Academy of Sciences, Beijing 100029, People’s Republic of China
4Department of Physics and Materials Science, City University of Hong Kong, Tat Chee Avenue, Kowloon, Hong Kong

(Received 5 August 2010; accepted 30 August 2010; published online 30 September 2010)

The origin of the flat band voltage roll-off ($V_{FB}$ roll-off) in metal gate/high-k/ultrathin-SiO2/Si metal-oxide-semiconductor stacks is analyzed and a model describing the role of the dipoles at the SiO2/Si interface on the $V_{FB}$ sharp roll-off is proposed. The $V_{FB}$ sharp roll-off appears when the thickness of the SiO2 interlayer diminishes to below the oxygen diffusion depth. The results derived using our model agree well with experimental data and provide insights to the mechanism of the $V_{FB}$ sharp roll-off. © 2010 American Institute of Physics. [doi:10.1063/1.3491292]

As a result of aggressive downscaling of metal-oxide-semiconductor (MOS) field-effect transistors metal and high-k oxides have replaced polycrystalline silicon and SiO2 as the gate materials. High-k oxides presently still underperform in comparison with SiO2 in some aspects. In order to improve the overall electrical properties of the metal gate/high-k/Si MOS stack, it is necessary to introduce an interlayer (IL) of SiO2 between the high-k dielectrics and silicon.1 If the thickness of the IL SiO2 is thinner, it is easier to obtain small equivalent oxide thickness (EOT). Unfortunately, when the IL SiO2 becomes ultrathin (~2 nm), a negative flat band shift called $V_{FB}$ Roll-off emerges2 and it is a big challenge to simultaneously achieve low threshold voltage ($V_{th}$) and thin EOT, especially in p-channel MOS devices. Recent reports have attempted to explain this phenomenon by bulk charge generation3,4 or bottom interface dipole at high-k/IL-SiO2 interfacial layer.5,6 However, the observed $V_{FB}$ roll-off is independent of EOT (Ref. 6) and it is contradictory to bulk charge generation. Furthermore, the model describing the dipole at the high-k/IL-SiO2 interface fails to explain the $V_{FB}$ sharp roll-off when the thickness of SiO2 becomes ultrathin.

In this work, the origin of the $V_{FB}$ sharp roll-off in metal gate/high-k/ultrathin-SiO2/Si MOS stacks is analyzed and the relationship between the dipoles at the IL SiO2/Si interface and the $V_{FB}$ sharp roll-off is studied theoretically. Our results indicate that when the thickness of the IL SiO2 is reduced below the O diffusion depth, the $V_{FB}$ roll-off appears abruptly. The results derived from this model are consistent with experimental data and illustrate that the $V_{FB}$ roll-off on ultrathin EOT can be predicted.

Generally, if $V_{FB}$ roll-off is induced by the dipole at the high-k/IL-SiO2 interface, the properties of the high-k layer should be one of the deciding factors for the $V_{FB}$ roll-off. However, similar to the HfO2/IL-SiO2/Si MOS stacks, the trend concerning $V_{FB}$ roll-off is also observed from Al2O3/IL-SiO2/Si stacks.6 Furthermore, when Al is incorporated into the HfO2 layer, an obvious positive shift in $V_{FB}$ (denoted as $V_{FB}$ roll-up) is observed prior to $V_{FB}$ roll-off.7 This indicates that Al doping can modulate the dipole located at the high-k/SiO2 interface and the $V_{FB}$ roll-off still remains. It means that the $V_{FB}$ roll-off can be influenced by other factors. Actually, the chemical bonds at the thermally grown SiO2/Si interfaces are easily converted to Si—O in which Si atoms have partial positive charges and O atoms are partially negatively charged.8 Thus, dipoles with a moment of about 3.1 D (1 D = 3.336 × 1023 C m) can be formed9 and it induces negative $V_{FB}$ shifts. Although it is well known that the magnitude of the Si—O dipoles depends mainly on the oxygen concentration, in the high-k/IL-SiO2/Si MOS stacks, O exists in O2− species and diffuses via exchange with lattice oxygen sites in the high-k layer.10 On account of the covalent Si—O bonds at the high-k/SiO2 interface, O diffusion from the high-k layer into IL SiO2 is suppressed11 and the concentration decreases rapidly with diffusion depth in the IL SiO2.12 Usually, there are two stages in the oxidation of silicon in the presence of the high-k layer. In the first stage for about 10 s, there is rapid growth to 1.5–3.0 nm which changes with annealing temperature.13,14 This is followed by slow growth in the apparent self-limited growth regime. In this stage, the thickness of the IL SiO2 contributes little even for long processing time at high annealing temperature.13,15 Hence, diffusing O is distributed over an ultrathin depth (1.5–3.0 nm) in the IL SiO2 which is close to the thickness showing sharp flat band voltage shifts. In order to quantitatively describe the concentration of diffusing O in the IL SiO2 and determine the magnitude of the dipole at the IL SiO2/Si interface, Fick’s diffusion law is used as follows: 

$$C = \frac{Q}{\sqrt{\pi Dt}} \exp \left( -\frac{r^2}{4Dt} \right),$$

(1)

∗Author to whom correspondence should be addressed. Electronic mail: ahuang@buaa.edu.cn. Tel.: 86-10-82317935, FAX: 86-10-82317935.
where $C$ is the concentration of O in the IL SiO$_2$, $Q$ is the diffusion activation energy, $D$ is the diffusion index (nm$^2$/s), $t_{in}$ is the diffused depth (nm), $t$ is the diffusion time (s), $D_0$ is the rate of diffusion, $K = 1.38 \times 10^{-23}$ J K$^{-1}$ is the Boltzmann constant, and $T$ is the diffusion temperature (°C). According to the diffusion index, the depth of diffusing O can be calculated by the following equation:

$$t_{in} = \lambda (D \times t)^{1/2} = \lambda \left[D_0 \exp\left(-\frac{Q}{KT}\right)\right]^{1/2},$$

where $\lambda$ is a constant.

When the thickness of IL SiO$_2$ is larger than the O diffusion depth, there is not enough dissociative O to form the interfacial dipoles between the IL SiO$_2$ and Si substrate. As the thickness of IL SiO$_2$ diminishes below the diffusion depth, Si=O dipoles form more easily and they impose negative sheet charges on the SiO$_2$ side and positive sheet charges on the Si substrate side, as illustrated in Fig. 1. The difference of potential $\phi_D$ can be calculated as follows:

$$\phi_D = E_0 d_{os} = \frac{Q_d \times d_{os} \times C_5}{\varepsilon_{int}} \times \frac{p \times C_S}{\varepsilon_{int}},$$

where $E_0$ is the electric field induced by the dipoles, $d_{os}$ is the distance between the two atoms which is equal to the Si=O bond length and approximately 1.52 Å, $Q_d$ is the electric quantity of the dipole, $C_5$ is the areal density of the dipoles, $\varepsilon_{int}$ is the dielectric constant of the SiO$_2$/Si interfacial layer, and $p$ is the dipole moment. In the metal gate/high-$k$/ultrathin-SiO$_2$/Si gate stacks, $V_{FB}$ can be expressed by the following equation considering the dipoles at the SiO$_2$/Si interface:

$$V_{FB} = \phi_{MS} - \phi_D = \frac{Q_1 \times T_{high-k}}{\varepsilon_0} - \frac{Q_2 \times EOT}{\varepsilon_ox},$$

where $\phi_{MS}$ is the effective work function of metal gate, $Q_1$ and $Q_2$ are charges (cm$^{-2}$) at the interface between high-$k$ and bottom oxide, and at the interface between bottom oxide and substrate, respectively, and $T_{high-k}$ is the physical thickness of the high-$k$ layer. Since the observed $V_{FB}$ roll-off is almost independent of the thickness of the high-$k$ layer, the influence of the interfacial charges shown in Eq. (5) can be ignored. Combining Eqs. (1), (2), and (4), $\Delta V_{FB}$ can be calculated as follows:

$$\Delta V_{FB} = \alpha \exp\left(-\frac{t_{in}}{\beta}\right),$$

where $\alpha = n(p/\varepsilon_{int})[Q/\sqrt{\pi D_{int} \exp(-Q/\varepsilon_{int})}]$ and $\beta = 4D_{int} \exp(-Q/\varepsilon_{int})$. In Eq. (7), it can be seen that $\Delta V_{FB}$ varies only as the $t_{in}$. When $t_{in}$ decreases below a specific value, $\Delta V_{FB}$ begins to increase rapidly. Therefore, in the curve of $V_{FB}$-EOT, the $V_{FB}$ sharp roll-off appears.

The experimental data of $\Delta V_{FB}$ obtained from the metal gate/HfO$_2$/beveled-SiO$_2$/Si gate stacks annealed at 400 °C (Ref. 16) are fitted by Eq. (7) and $\alpha = 800$ and $\beta = 25$ are obtained as shown in Fig. 3. A value of $V_{FB} \sim 800$ mV is obtained when the thickness of IL SiO$_2$ approaches zero and it agrees well experimental data (700–800 mV). According to Fick’s law, the diffusion depth increases with annealing temperature. Compared to the thickness of about 1.4 nm in Fig. 3, when the metal gate/HfO$_2$/beveled-SiO$_2$/Si MOS stack is postannealing at 1000 °C, the $V_{FB}$ sharp roll-off is observed at an IL SiO$_2$ thickness of about 3.5 nm. The constant $\beta$ can be determined from with Eqs. (3) and (7) as follows:

$$\lambda t_2 = \left[\frac{D_0 \exp\left(-\frac{Q_2}{KT_2}\right)}{t}\right]^{1/2} = \sqrt{\frac{1}{\beta_2}} = \frac{1.4}{3.5} \Rightarrow \beta_2 = 156.25.$$
the same magnitude of the $V_{FB}$ shift when HfO$_2$ contacts Si directly.\cite{17} The equation of $\Delta V_{FB}$ at an annealing temperature of 1000 °C is obtained by the following:

$$\Delta V_{FB} = 750 \exp \left( \frac{-t_n}{156.25} \right). \tag{9}$$

Figure 4 shows the $\Delta V_{FB}$ of the metal gate/HfO$_2$/beveled-SiO$_2$/Si MOS stack after postannealing at 1000 °C and the curve of Eq. (9). The results clearly reveal that when the thickness of IL SiO$_2$ is below the O diffusion depth, $\Delta V_{FB}$ increases sharply as verified by experiments. As shown in Fig. 3, a low annealing temperature reduces $\Delta V_{FB}$ for the same thickness of IL SiO$_2$. It means that O incorporation at low temperature is an effective method to reduce the $V_{FB}$ sharp roll-off without increasing EOT.\cite{14} Besides, it is helpful to decrease the $V_{FB}$ sharp roll-off for post-annealing without O atmosphere.\cite{19} On the other hand, when the top SiO$_2$ layer between the metal gate and high-$k$ is introduced, the Si–O dipoles form due to oxidation at the metal gate/top-SiO$_2$ interface. It has the opposite dipole moment as the dipole at IL SiO$_2$/Si interface and so $\Delta V_{FB}$ is reduced. However, when the thickness of the top-SiO$_2$ layer is larger than the depth of diffusing O, the top dipoles cannot form and the $V_{FB}$ roll-off reappears. Hence, the $V_{FB}$ sharp roll-off can also be eliminated by inserting 1–2 nm of SiO$_2$ between the metal gate and high-$k$ layer as experimentally verified.\cite{20}

In conclusion, the origin of the $V_{FB}$ sharp roll-off in the metal gate/high-$k$/ultrathin-SiO$_2$/Si MOS stacks has been investigated. An excessive amount of dissociative O diffuses into the IL SiO$_2$ during process and forms O–Si dipoles at the ultrathin-IL SiO$_2$/Si interface consequently playing an important role in the $V_{FB}$ sharp roll-off. An extra voltage induced by the dipoles shift the flat band voltage negatively. The results derived from our model well agree with the experimental data and provide insight to the mechanism associated with the $V_{FB}$ sharp roll-off.

Our work was financially supported by National Natural Science Foundation of China (Grant Nos. 50802005 and 11074020), Program for New Century Excellent Talents in University (Grant No. NCET-08-0029), Ph.D. Program Foundation of Ministry of Education of China (Grant No. 20080061055), and Hong Kong Research Grants Council (RGC) General Research Funds (GRF) Grant No. CityU 112608.