Determination of capture cross sections for as-grown electron traps in Hf O 2 ∕ Hf Si O stacks

C. Z. Zhao, J. F. Zhang, M. B. Zahid, B. Govoreanu, G. Groeseneken, and S. De Gendt

Articles you may be interested in

Charge trapping defects in Si / SiO 2 / Hf (1 − x) Si x O 2 film stacks characterized by spectroscopic second-harmonic generation
J. Vac. Sci. Technol. B **29**, 04D101 (2011); 10.1116/1.3591433

Photoinduced charge-trapping phenomena in metal/high- k gate stack structures studied by synchrotron radiation photoemission spectroscopy
Appl. Phys. Lett. **96**, 162902 (2010); 10.1063/1.3409162

Charge trapping properties in Ti O 2 ∕ Hf Si O ∕ Si O 2 gate stacks probed by scanning capacitance microscopy
Appl. Phys. Lett. **92**, 012112 (2008); 10.1063/1.2828863

Charge trapping and dielectric relaxation in connection with breakdown of high- k gate dielectric stacks
Appl. Phys. Lett. **88**, 202904 (2006); 10.1063/1.2203942

Spatial distributions of trapping centers in HfO 2 ∕ SiO 2 gate stacks
Appl. Phys. Lett. **88**, 152907 (2006); 10.1063/1.2195896

Determination of capture cross sections for as-grown electron traps in HfO$_2$/HfSiO stacks

C. Z. Zhao, J. F. Zhang,[a)] and M. B. Zahid
School of Engineering, Liverpool John Moores University, Byrom Street, Liverpool L3 3AF, United Kingdom

B. Govoreanu, G. Groeseneken,[b)] and S. De Gendt[b)]
IMEC, Kapeldreef 75, B3001 Leuven, Belgium

A major challenge for replacing gate SiON with HfO$_2$ is the instability and reliability of HfO$_2$. Unlike the SiON, there can be substantial amount of as-grown electron traps in HfO$_2$. These traps can cause instability in the threshold voltage and contribute to the dielectric breakdown. Despite the early efforts, our understanding of them is incomplete. Agreement on their capture cross sections has not been reached and the reported values spread in a large range of 10^{-12}–10^{-19} cm^2. The objective of this paper is to determine their capture cross sections unambiguously, which requires knowing the gate current and the electron fluency for filling the trap. A key part of this work is to estimate the trapping-induced transient gate current following the application of a pulse to the gate. This is achieved by numerical simulation. It is found that trapping can reduce the gate current by two orders of magnitude and the gate current can drop substantially within microseconds. The results show the presence of two distinctive capture cross sections in the order of 10^{-14} and 10^{-16} cm^2, respectively, which most likely originated from two different types of as-grown electron traps in HfO$_2$. These capture cross sections are insensitive to fabrication and processing techniques. © *2006 American Institute of Physics.* [DOI: 10.1063/1.2364043]

I. INTRODUCTION

A major challenge for the current complementary metal-oxide-semiconductor (CMOS) industry is the rapid increase of gate leakage current as the thickness of SiON approaches 1 nm. To reduce this leakage, intensive worldwide efforts have been made to find an alternative gate dielectric of higher dielectric constant (high k) than SiON.[1–10] For the same equivalent electrical oxide thickness (EOT), high-k layers are physically thicker and it has been demonstrated that gate leakage can be reduced by several orders of magnitude when Hf-based dielectric was used.[1] What holds back the commercial application of these high-k layers includes threshold voltage V_{th}, instabilities, gate work function instabilities, gate–high-k layer interaction, lower carrier mobility, and gate dielectric breakdown.[1–10]

Although as-grown electron traps are negligible in a modern SiON, it has been found that they exist in large numbers ($\sim 10^{13}$ cm^{-2}) in HfO$_2$ layers.[4,5] These traps are generally believed to be responsible for V_{th} instability[1–5] and they also contribute to the breakdown.[6] Understanding these traps is of importance. Early work[7] showed that these traps have an energy band of 0.5–0.8 eV below the bottom edge of HfO$_2$ conduction band. The electron trapping is highly dynamic and trap levels can be substantially underestimated when measured by conventional methods based on the recording of the shift of transistor's dc transfer characteristics.[1,4,8,10] Despite the early efforts, many issues remain to be resolved.

One main issue is the determination of their capture cross sections. The capture cross section σ is one of the most important properties for traps. Agreement has not been reached on whether σ has discrete or continuously distributed values.[3–5,9] The reported capture cross sections spread over a range as large as 10^{-12}–10^{-19} cm^2.[3–5,9,10] At the large end of this range, $\sigma = 3 \times 10^{-13}$ cm^2 was obtained based on irradiation experiments, where the electron fluency was limited to 3×10^{13} cm^{-2}.[9] This is too low for detecting traps of smaller cross sections.[11] At the small end of this range, capture cross sections in the order of 10^{-19} cm^2 was obtained from the shift of quasi-dc transfer characteristics.[3] In this case, we will show that the significant detrapping has a material effect on the extraction of capture cross section.

The objective of this work is to determine the capture cross section of as-grown electron traps in HfO$_2$ unambiguously. The test conditions are fine tuned for this task. Detrapping will be suppressed and the range of electron fluency has been selected so that trapping is negligible at the start, but saturates at the end. To determine the electron fluency through the gate dielectric, gate current is needed. A key part of this work is to estimate the transient gate current when a pulse is applied to the gate. This is achieved through numerical modeling. It will be shown that trapping can reduce the gate current substantially in microseconds. The electron fluency can be underestimated by one order of magnitude when this transient of gate current is not taken into account. Our results support the presence of two well separated capture cross sections rather than a continuous distribution. The extracted values are in the order of 10^{-14} and 10^{-16} cm^2. The traps with capture cross sections of 10^{-14} and 10^{-16} cm^2 will

[a)]Electronic mail: j.f.zhang@ljmu.ac.uk
[b)]Also at KU Leuven, Leuven, Belgium.
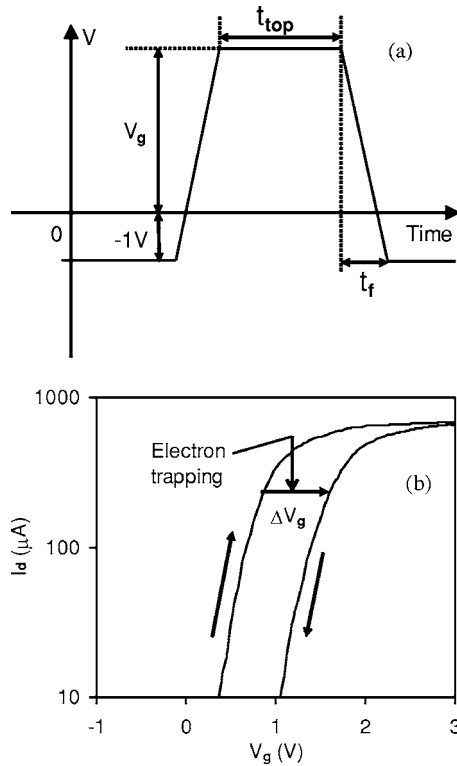
100, 093716-1

FIG. 1. The pulse applied to the gate (a) and the pulsed transfer characteristics (b). A -1 V was first applied to the gate to empty any trapped electrons. During the rising edge of the pulse, trapping is negligible and the corresponding $I_d \sim V_g$ is the left curve in (b). Traps were filled during the period of t_{top}. This leads to a shift of gate bias, ΔV_g, when the $I_d \sim V_g$ was measured again during the falling edge of the pulse. The delay between trap filling and measurement is less than the falling time, t_f.
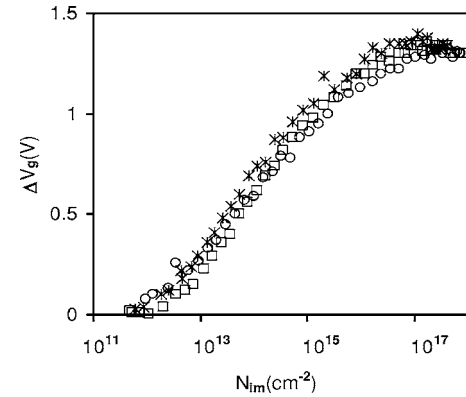

FIG. 2. Typical results of as-grown electron trapping. The electron fluency, N_{im}, was obtained from Eq. (1) based on the gate current density given in Fig. 3. Each point corresponds to one ΔV_g in Fig. 1(b) and was obtained by applying one pulse to the gate. The t_{top} and t_f were fixed at 23 and 10 μs, respectively. To increase N_{im}, the gate bias amplitude was increased progressively by a step of 0.1 V, starting from $V_g = 1.6$ V. The same tests were carried out on three devices and the results are represented by the three sets of symbols. The sample-to-sample variations are insignificant.

be referred to as "large" and "small" traps, respectively, hereafter. For a 4 nm HfO$_2$ prepared by atomic layer deposition (ALD), the effective densities for the large and small traps are 5.3×10^{12} and 1.1×10^{13} cm^{-2}, respectively.

II. DEVICES AND EXPERIMENT

A. Devices

To assess the capture cross section of electron traps, it is desirable to select a test sample with a large amount of as-grown traps. Our early work[4] shows that the effective density of as-grown electron traps in a 4 nm ALD HfO$_2$ is in the order of 10^{13} cm^{-2}. The 4 nm ALD HfO$_2$ was chosen as the test sample in this work, therefore. The physical vapor deposited (PVD) metal gate consists of a 10 nm TaN layer capped with a 70 nm TiN. Before HfO$_2$ deposition, there was a 0.4 nm chemical oxide. To activate the dopant, a 1000 °C and 1 s anneal was used. The transmission electron microscopy (TEM) analysis shows that there is an interfacial layer (IL) of 1.7 nm, which contains Hf silicate.[12] At the end of the processing, a forming gas anneal was carried out at 520 °C for 20 min. The EOT of the stack is 1.75 nm. The size of n-type metal-oxide-semiconductor field-effect transistors (n-MOSFETs) used is 0.25 μm in length and 10 μm in width.

To assess the impact of gate materials and related processes on electron trapping, limited tests were carried out on poly-Si gated MOSFETs with the same HfO$_2$ thickness. The

poly-Si gated devices have an initial 1 nm chemical oxide and the dopant activation was at 1000 °C for 10 s, resulting in an EOT of 1.8 nm. To further explore the effect of fabrication techniques on electron trapping, a 2 nm HfO$_2$ was also prepared by PVD. The oxidation was at 600 °C for 5 s, followed by a postdeposition anneal at 700 °C in N$_2$ for 60 s.

B. Measurements

As mentioned earlier, to extract the capture cross section, detrapping should be suppressed to simplify the test condition. This requires using the pulsed $I_d \sim V_g$ technique.[8] Figure 1(a) shows the typical pulse applied to the gate. A -1 V was first applied for a sufficiently long time (~ 1 s) to "reset" the sample by emptying traps through detrapping. During the rising edge of the pulse, drain current was monitored to give the $I_d \sim V_g$ on the left in Fig. 1(b). Traps were filled with electrons during the period of t_{top}. The negative trapped charges lead to a positive shift of the $I_d \sim V_g$ during the falling edge of the pulse, as shown by the curve on the right in Fig. 1(b). The trapping level is measured from this shift, ΔV_g. The measurement time is within the falling period, t_f, and its impact on the result will be evaluated in Sec. II C.

A typical result is given in Fig. 2. Each point in Fig. 2 was obtained by applying one pulse as shown in Figs. 1(a) and 1(b). To increase the electron number injected into HfO$_2$ for filling traps, the gate bias at the t_{top}, V_g, was progressively increased in a step of 0.1 V. Figure 2 shows that electron fluency in the order of 10^{17} cm^{-2} is needed for trapping to reach saturation. The electron fluency, N_{im}, at a given V_g in Fig. 2 is calculated from

$$N_{im} = J_{gm} \times t_{top}/q + \Delta V_g \times C_{ox}/q, \qquad (1)$$

where J_{gm} is the measured gate current per unit area, q is one electron charge, and C_{ox} is the gate oxide capacitance per unit area. It should be pointed out that the N_{im} in Eq. (1) is the fluency of electrons injected into the gate dielectric rather than exiting from the dielectric. The inclusion of the second
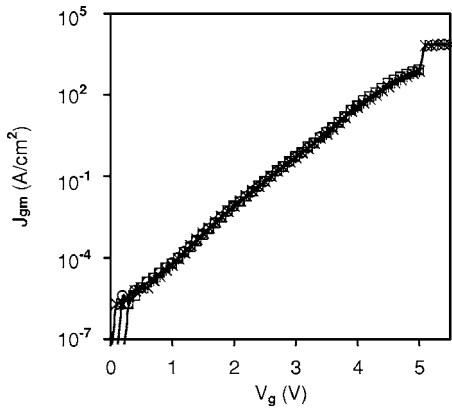
FIG. 3. The measured gate current per unit area against gate biases. Here, the measurement of each point took approximately 0.15 s, which was much longer than the pulse period used in Fig. 2. The same measurements were made on three devices and the sample-to-sample variations were small.

FIG. 5. Dependence of trapping levels on the pulse falling time, t_f. The electron fluency was fixed at 10^{17} cm^{-2}. When t_f is higher than 30 μs, an increase of t_f enhances detrapping, which in turn reduces the trapping. For $t_f < 30$ μs, detrapping is not important and trapping is insensitive to t_f.

term in Eq. (1) is essential, since it makes substantial contribution at low N_{im}. The derivation of Eq. (1) is given in the Appendix.

The typical J_{gm} is plotted in Fig. 3. Figures 2 and 3 also show that the sample-to-sample variation is acceptable. As a result, the quality of the data is good enough for extracting electron capture cross sections. All measurements were at room temperature.

Figure 4 compares electron trapping in samples of different gates. Although trapping in metal-gated samples is lower, the impact of gate materials is modest. Since the industrial attention is focused on metal gates at present, we will also concentrate on metal-gated samples in this work.

C. Selection of measurement time

As mentioned in Sec. I, when the trapping in HfO$_2$ was assessed from the shift of quasi-dc transfer characteristics in the subthreshold region,[13] the measurement can take seconds and Fig. 4 shows that trapping is substantially underestimated through detrapping. If these quasi-dc data are used for
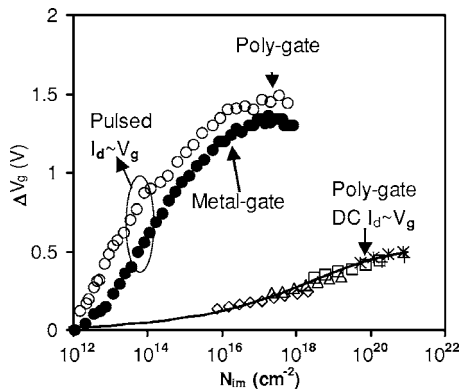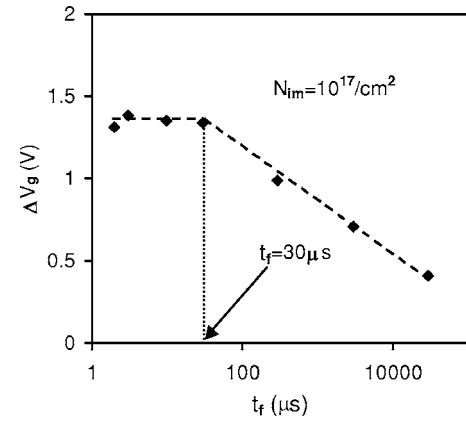

FIG. 4. Effects of gate materials and measurement techniques on electron trapping. Changing gate from poly-Si to metal leads to a modest reduction of trapping, but the overall features remain the same. The t_{top} and t_f for pulsed measurement are 23 and 10 μs, respectively. Trapping is significantly reduced through detrapping during the quasi-dc measurement. The solid line was obtained by fitting with Eq. (17). The gate voltage used for the quasi-dc measurement is "◇"—2 V, "△"—2.5 V, "□"—3 V, "*"—3.5 V, and "+"—4 V.

extracting capture cross sections, a value in the order of 10^{-19} cm^2 is obtained, which agrees well with that reported in early works[3,5] and indicates that capture cross section is not sensitive to the difference in samples used by different groups. However, a comparison with the pulsed data clearly shows that the detrapping during quasi-dc measurements leads to a severe underestimation of the capture cross section. The question is how fast the pulse has to be to suppress detrapping.

For the pulsed $I_d \sim V_g$ technique, Figs. 1(a) and 1(b) show that the delay between trap filling and measurement is controlled by the falling time t_f of the pulse. Tests were carried out to assess the impact of t_f on trapping. For a given N_{im}, Fig. 5 shows that trapping, indeed, reduces substantially for long t_f through detrapping. However, trapping is insensitive to t_f for $t_f < 30$ μs. This indicates that detrapping becomes insignificant when $t_f < 30$ μs for the sample used in this work. $t_f = 10$ μs is selected for the rest of this work.

III. ESTIMATION OF TRANSIENT GATE CURRENT

To determine the capture cross section, the electron fluency must be known. One way of obtaining the fluency is by measuring the transient gate current during the gate pulse and integrating it against time. Unfortunately, in practice, we could not measure this small and transient gate current, since the parameter analyzer is too slow. The J_{gm} measurement in Fig. 3 took about 0.15 s, while the typical transient is in the order of microseconds. Electron trapping reduces the electrical field near the cathode and, consequently, leads to a reduction of the gate current.[14] Figure 2 shows that trapping can induce a gate voltage shift over 1 V. An inspection of Fig. 3 shows that, for this amount of reduction in V_g, J_{gm} can be reduced by two orders of magnitude. As a result, there will be a large change in the gate current when a pulse is applied to the gate. Initially, there is no trapping and gate current will be high. As trapping accumulates, gate current will approach the J_{gm} shown in Fig. 3.

The N_{im} in Fig. 2 was calculated from the measured gate current density J_{gm} by using Eq. (1). Here, the assumption is

that the gate current is a constant during the pulse and is equal to its value shown in Fig. 3. This will lead to an underestimation of electron fluency, because the transient high current was not taken into account. In this section, efforts will be made to reduce the error in N_{im}. Since we cannot measure the transient gate current, we attempt to give a first order estimation through numerical simulation.

A. Numerical modeling of gate current without trapping

Under a given gate bias, electrons can tunnel through the thin dielectric stack with or without the assistance of traps.[15] The current caused by trap-assisted tunneling (TAT) will not be considered here, since our recent work[4] clearly shows that it does not contribute to trapping. Without TAT, tunneling current per unit area from the substrate, J_g, can be evaluated by[16–18]

$$J_g = q\sum_{ij} N_{ij}T_{ij}f_{ij}. \tag{2}$$

The simulation can be broadly divided into four steps. First, one must know the density of charge carriers available for tunneling, N_{ij}, from the inversion layer in the jth energy subband of the ith valley.[16–18] Second, the potential profile in the dielectric is determined, since tunneling probability T_{ij} is sensitive to it. T_{ij} is then calculated based on the Wentzel-Kramers-Brillouin (WKB) approximation. Finally, the impact frequency at the dielectric/substrate interface f_{ij} is calculated and J_g is obtained from Eq. (2). More information for each step is given below.

1. Calculation of N_{ij} and E_{ij}

To start the simulation, we chose the substrate surface potential V_s as the input parameter. The corresponding N_{ij} and the subband energy level E_{ij} were obtained by solving the following equations numerically:[16,17]

$$E_{ij} = \left(\frac{\hbar^2}{2m_{zi}}\right)^{1/3}(A_j q F_{eff})^{2/3} \tag{3}$$

and

$$N_{ij} = \left(\frac{kT}{\pi\hbar^2}\right)g_i m_{di}\ln\left[1 + \exp\left(\frac{E_F - E_{ij}}{kT}\right)\right], \tag{4}$$

where $\hbar = h/(2\pi)$ and h is Planck's constant, m_{zi} is the effective mass in the ith valley, F_{eff} is the effective electrical field strength, and A_j originates from the jth zero of the Airy function,

$$A_j \approx \frac{3}{2}\pi\left(j + \frac{3}{4}\right) \quad \text{with } j = 0,1,2,\ldots. \tag{5}$$

In Eq. (4), k is Boltzmann's constant, T is the temperature, m_{di} is the density-of-states effective mass, g_i is the degeneracy of the ith valley, and E_F is the Fermi level. The values of g_i, m_{zi}, and m_{di} used for the calculation are given in Table I.

TABLE I. Parameters used for the ith valley of Si conduction band. g_i, m_{zi}, and m_{di} are the degeneracy, effective mass of electrons, and the density-of-states effective mass of the ith valley. m_0 is the free electron mass. The data were taken from the references in the square bracket.

	g_i [Ref. 17]	m_{zi} [Ref. 17] (m_0)	m_{di} [Ref. 17] (m_0)
$i=0$	2	0.916	0.190
$i=1$	4	0.190	0.417

2. Determination of potential drop in the dielectric and gate voltage

Once the N_{ij} is known, the total carrier density in the inversion layer, N_{inv}, can be obtained by summing up N_{ij} in all subbands. The total potential drop over the oxide V_{ox}^0 is given by[16]

$$V_{ox}^0 = \frac{q(N_{depl} + N_{inv})\text{EOT}}{\varepsilon_{SiO_2}\varepsilon_0}, \tag{6}$$

where N_{depl} is the charged dopant per unit area in the depletion layer and the superscript "0" represents trapping-free. The EOT is evaluated by

$$\text{EOT} = t_{IL}\varepsilon_{SiO_2}/\varepsilon_{IL} + t_{HfO_2}\varepsilon_{SiO_2}/\varepsilon_{HfO_2}, \tag{7}$$

where ε_{SiO_2}, ε_{IL}, and ε_{HfO_2} are the dielectric constant of SiO$_2$, the interfacial layer (Hf silicates), and HfO$_2$, respectively. t_{HfO_2} and t_{IL} are the thickness of HfO$_2$ and the interfacial layer, and ε_0 is the permittivity of free space. The voltage drop over each layer of the stack (see Fig. 6) is determined from

$$V_{IL}^0 = \frac{V_{ox}^0}{\text{EOT}}\frac{t_{IL}\varepsilon_{SiO_2}}{\varepsilon_{IL}} \tag{8}$$

and

$$V_{HfO_2}^0 = V_{ox}^0 - V_{IL}^0. \tag{9}$$

The gate voltage is evaluated by

$$V_g^0 = V_s + V_{ox}^0 + \psi_{ms}, \tag{10}$$

where ψ_{ms} is the work function difference.

3. Evaluation of T_{WKBij}

Once the potential drop over the dielectric is known, the tunneling probability can be evaluated. As a first order estimation, T_{ij} is evaluated based on the WKB approximation[18]

$$T_{ij} = T_{WKBij}T_{Rij}, \tag{11}$$

where T_{Rij} is a correction factor, taking into account reflections at interfaces due to energy band discontinuities, and its calculation can be found in Ref. 18. The WKB tunneling probability T_{WKBij} can be expressed as[19]

$$T_{WKBij} = \exp\left[\frac{-2\sqrt{2m_{ox}}}{\hbar}\int_0^x\sqrt{\Phi(x) - E_{ij}}\,dx\right], \tag{12}$$

where m_{ox} is the effective mass of tunneling electrons in the oxide and $\Phi(x)$ is the bottom edge of conduction band in the
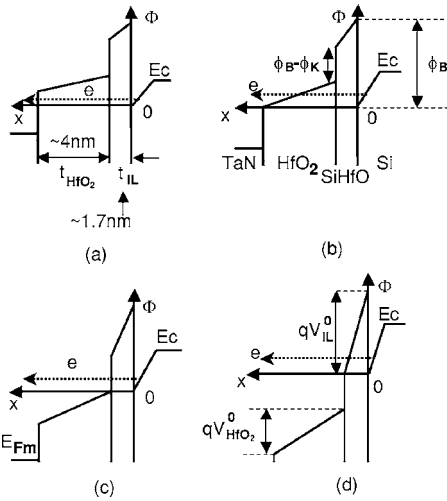
FIG. 6. Schematic energy band diagrams for different tunneling regimes. The gate bias increases from case (a) to case (d). (a) shows direct tunneling through both HfO$_2$ and interfacial layers (ILs). (b) is the case where electrons emerge from the conduction band of HfO$_2$. (c) illustrates that electrons only tunnel through IL directly. Finally, Fowler-Nordheim (FN) tunneling occurs through the IL in (d). ϕ_k in (b) is the conduction band offset between HfO$_2$ and silicon.

oxide. As shown in Figs. 6(a)–6(d) there are four possible cases for electron tunneling. Figure 6(a) shows direct tunneling through both dielectric layers at low V_g. An increase of V_g leads to electron emerging from the conduction band of HfO$_2$ [Fig. 6(b)]. Figure 6(c) illustrates that electrons only tunnel through the interfacial layer directly at higher V_g. Finally, Fowler-Nordheim tunneling through the IL occurs [Fig. 6(d)].

4. Calculation of J_g

The impact frequency at the dielectric/substrate interface in Eq. (2) is calculated from[16,17]

$$f_{ij} = \frac{E_{ij}}{2A_j\hbar}. \qquad (13)$$

Finally, J_g is obtained from Eq. (2). The gate current I_g is obtained from the product of J_g with gate area. We found that I_g is dominated by the first subband in the two valleys. Higher subbands have lower N_{ij} and contribute little to I_g.

The result is presented in Fig. 7. The top curve is for the trapping-free case, while the cases with trapping will be described in the next section. The parameters used for the simulation are given in Tables I and II. It should be noted that we have a Hf-silicate interfacial layer[12] and the energy barrier height at the IL/substrate interface, ϕ_B, and the effective electron mass, m_{IL}, were not known. ϕ_B was set at 2.35 eV and m_{IL} at $0.45m_0$, so that the calculated I_g agrees with the measured I_g when trapping is insignificant at V_g = 2 V. m_0 is the free electron mass. Since ϕ_B=3.15 eV for SiO$_2$ and 1.5 eV for HfO$_2$,[17] ϕ_B=2.35 eV is reasonable for Hf silicates. m_{IL}=$0.45m_0$ is also acceptable for Hf silicates, because of m_{HfO_2}=$0.18m_0$ and m_{SiO_2}=$0.5m_0$.[17]

The voltage range corresponding to each tunneling case illustrated in Figs. 6(b)–6(d) is marked out on Fig. 7. The direct tunneling through both layers [Fig. 6(a)] cannot be
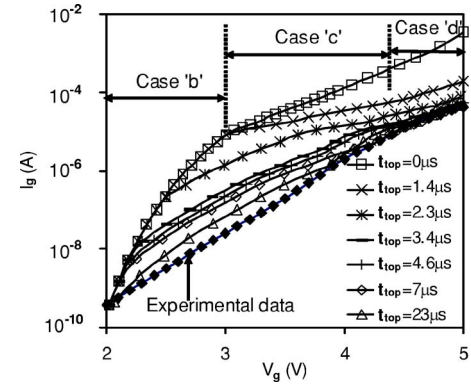

FIG. 7. A comparison of the measured and calculated gate currents with various pulse times, t_{top}. The solid curves are guides for the eyes. The top curve (t_{top}=0) represents the trapping-free case. The range of V_g corresponding to each tunneling regime of Figs. 6(b)–6(d) is marked out for t_{top}=0. The case corresponding to Fig. 6(a) is not shown, since it occurs at V_g<2 V. An increase of either V_g or t_{top} leads to a decrease of I_g from its trapping free level and it is clear that trapping can reduce I_g substantially. At high V_g and t_{top}, trapping saturates and, consequently, the difference between the measured and calculated I_g disappears and I_g becomes insensitive to t_{top}.

seen from Fig. 7, since it occurs at a voltage less than 2 V. When electrons enter into the conduction band in HfO$_2$ [Fig. 6(b)], the current rises rapidly with voltage. As V_g reaches 3 V approximately, electrons only directly tunnel through the IL [Fig. 6(c)]. Since the tunneling distance does not decrease for higher V_g in this case, I_g rises relatively slowly, which explains the smaller $\log(I_g) \sim V_g$ slope in case c shown in Fig. 7. Finally, Fowler-Nordheim (FN) tunneling takes place through the IL when V_g>4.4 V.

Figure 7 clearly shows that the calculated I_g, assuming trapping-free, can be two orders of magnitude higher than the measured I_g, where trapping took place. This confirms our expectation and we will attempt to estimate I_g in the presence of trapping next.

B. Numerical modeling of gate current with trapping

Although many articles were published on the calculation of electron tunneling through high-k/IL stacks,[16,17,20–24] there is hardly any work that was carried out in the presence of transient trapping. When trapping occurs, the potential distribution profile in the dielectric changes with time. The main difficulty is how to estimate this transient distribution. It is overcome here by experimentally measuring the ΔV_g induced by trapping and some typical results are given in Fig. 8. Once the ΔV_g at a given bias and time is known, a first order estimation of the potential distribution and the corresponding gate current can be made, as described below.

To determine the potential distribution in the dielectric for a given ΔV_g, the spatial distribution of trapped charges is required. Unfortunately, this information is not available at present. Since it is known that as-grown electron traps are negligible in SiO$_2$ (Refs. 13 and 25) and Hf silicate,[26] the traps must be located either at the HfO$_2$/IL interface or in the bulk of HfO$_2$. By examining how trapping depends on the HfO$_2$ thickness, we can rule out that traps are concentrated at the HfO$_2$/IL interface.[27,28] Early work[10] also indicates that

TABLE II. Parameters used for calculating gate current through the HfO$_2$/HfSiO stack. ψ_{ms} is the work function difference. ϕ_B is the conduction band offset between the interfacial layer and silicon. ϕ_k is the conduction band offset between HfO$_2$ and silicon. N_A is the substrate doping density. In Eq. (12), the effective electron mass m_{ox} is replaced by m_{HfO_2} and m_{IL} in the HfO$_2$ and interfacial layer, respectively. The data were taken from the references in the square bracket. $\varepsilon_{IL} = 6.84$ was obtained from Eq. (7) to give an EOT of 1.75 nm. ϕ_B and m_{IL} were selected so that the calculated I_g agrees with the measured value when trapping is insignificant at $V_g = 2$ V.

ψ_{ms} [Ref. 12] (V)	ε_{IL}	ε_{HfO_2} [Ref. 12]	t_{IL} [Ref. 12] (nm)	t_{HfO_2} [Ref. 12] (nm)	EOT [Ref. 12] (nm)	ϕ_B (eV)	ϕ_k [Ref. 16] (eV)	m_{IL} (m_0)	m_{HfO_2} [Ref. 21] (m_0)	N_A [Ref. 17 and 18] (cm^{-3})
-0.55	6.84	20	1.7	4	1.75	2.35	1.5	0.45	0.18	5×10^{17}

trapping occurs in the bulk of HfO$_2$. As a first order estimation, we assume that the centroid of trapped charges is at the middle of HfO$_2$ layer.

The calculation follows the same procedure as that without trapping, apart from the evaluation of potential drop in the dielectric. Here, the presence of trapped charges must be taken into account. As shown in Fig. 9, for a given substrate surface potential V_s, trapping in HfO$_2$ does not affect the potential drop over the IL, but increases the voltage drop over HfO$_2$. The total potential drop over the stack is now

$$V_g = V_g^0 + \Delta V_g. \tag{14}$$

When there is no trapping, V_g^0 is fixed for a given V_s. Since trapping increases with time, ΔV_g and V_g will be a function of time for a given V_s now. As a result, to find a definitive ΔV_g and V_g, one has to fix both V_s and time.

As expected, Fig. 8 shows that an increase of either gate voltage or time (t_{top}) results in higher trapping. Once V_s and t_{top} are fixed, V_g^0 is known and we can set ΔV_g at a guessed initial value. This will give us two V_g: one from Eq. (14) and the other from the relevant curve in Fig. 8. If these two V_g do not agree with each other, ΔV_g will be adjusted until their difference becomes negligible.

Figure 7 shows the calculated I_g in the presence of trapping. For a given t_{top}, at relatively lower V_g, trapping is negligible and I_g agrees with the trapping-free value. An increase of V_g enhances trapping and brings down I_g. For a given V_g, an increase of t_{top} leads to higher trapping and lower I_g. It is important to note that when V_g and t_{top} are sufficiently high, the calculated I_g agrees with the measured value and becomes insensitive to t_{top}. This is because trapping reaches

saturation now, as shown in Fig. 2, so that I_g becomes insensitive to t_{top}. The reasonable agreement between the measured and calculated I_g at saturation makes us believe that the calculated I_g is acceptable as a first order approximation.

IV. EXTRACTION OF CAPTURE CROSS SECTIONS

A. Estimation of electron fluency

To determine the capture cross section of electron traps, electron fluency must be known. From Fig. 7, the transient gate current for a given V_g can be obtained. One example is shown in Fig. 10 for $V_g = 3.5$ V. As expected, I_g can drop by two orders of magnitude before reaching the measured value. When electron fluency N_{im} was calculated from Eq. (1) based on the measured gate current, its first term is schematically represented by the gray region in Fig. 10. By integrating the transient I_g against time, a more accurate electron fluency N_{ic} can be obtained by

$$N_{ic} = \frac{\int_0^{t_{top}} I_g dt}{qLW} + \frac{\Delta V_g \varepsilon_{HfO_2} \varepsilon_0}{t_{HfO_2}/2}, \tag{15}$$

where L and W are channel length and width, respectively. The first term of Eq. (15) is schematically represented by the striped region in Fig. 10.

In Fig. 11, the trapping-induced ΔV_g is plotted against both N_{im} and N_{ic}. At low fluency, trapping is negligible and there is little difference in these two. As the fluency and trapping increase, N_{ic} can be over one order of magnitude
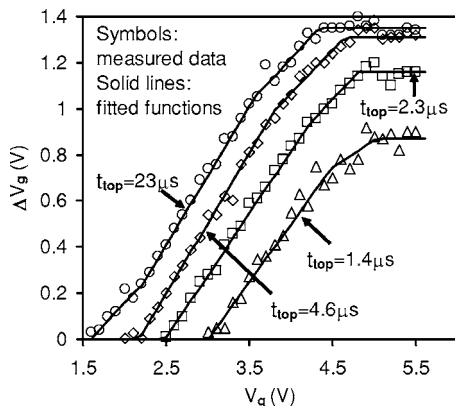
FIG. 8. The electron trapping when different gate voltages were applied for a time of t_{top}. The solid lines are fitted functions. An increase of either voltage or t_{top} enhances trapping.
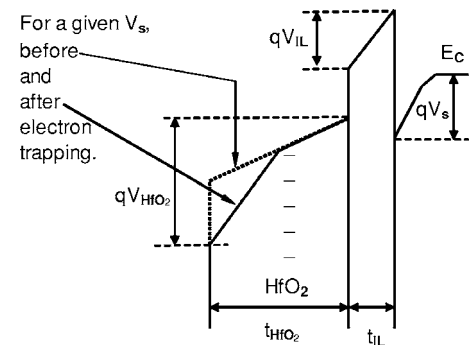
FIG. 9. Schematic energy band diagram showing the effect of electron trapping on potential distribution for a given substrate surface potential. The charge centroid is assumed to be at the center of HfO$_2$. The dotted and solid lines represent the cases without and with electron trapping, respectively. Trapping leads to an increase in the electrical field near the gate.
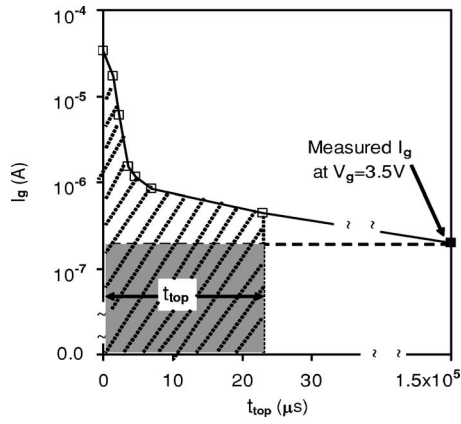
FIG. 10. The calculated transient gate current when $V_g = 3.5$ V was applied to the gate. For comparison, the measured I_g is also shown. Trapping reduces I_g by two orders of magnitude and the substantial reduction occurs within microseconds. The gray area schematically represents the first term of Eq. (1), where the measured gate current was used to determine the electron fluency and the transient of I_g was not taken into account. The striped area schematically shows the first term of Eq. (15), where the transient gate current was used to calculate electron fluency. Note the logarithmic scale for the current.
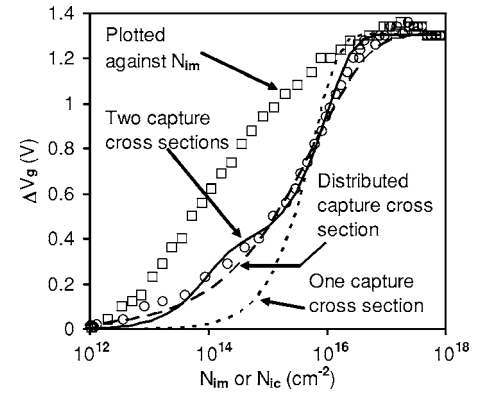

FIG. 11. A comparison of electron fluency calculated with and without taking the transient of gate current into account. For the symbol "□" and "○," the same data were used for the vertical axis, ΔV_g, but different data were used for the horizontal axis. The symbol □ was plotted against N_{im}, calculated from Eq. (1) without taking the transient of gate current into account. The symbol ○ was plotted against N_{ic}, calculated from Eq. (15) where the transient gate current is used. It shows that N_{ic} can be one order of magnitude higher than N_{im}. The dotted and solid lines were obtained by fitting the data with Eq. (16) when using one and two discrete capture cross sections, respectively. The dashed line is obtained by fitting the data with Eq. (17), assuming there is a distribution of capture cross sections.

larger than N_{im}. In the saturation region, the relative difference between N_{im} and N_{ic} decreases because the transient period of I_g becomes relatively less important.

B. Extraction of capture cross sections

We would like to assess the range of capture cross sections first from a direct observation of the data in Fig. 11. Considerable trapping occurs before electron fluency reaches 10^{14} cm^{-2}, indicating that there are traps with an effective capture cross section in the order of 10^{-14} cm^2. There is little further trapping for $N_{ic} > 10^{17}$ cm^{-2}. This means that there are no as-grown traps for a capture cross section of less than 10^{-17} cm^2. To extract more definitive values for capture cross sections, a trapping model is needed.

As described in the Sec. II C, we have selected test conditions to ensure that detrapping is suppressed. In this case, there are two popular trapping models. One is the first order model with discrete capture cross sections[11,25]

$$\Delta V_g = \sum_{k=1}^{K} \Delta V_{gk}(1 - e^{-\sigma_k N_{ic}}),$$ (16)

where K is the number of discrete capture cross sections, σ_k. The other assumes that the value of capture cross section has a distribution[3]

$$\Delta V_g = V_0[1 - e^{-(\sigma_0 N_{ic})^\beta}].$$ (17)

If we use Eq. (16) with a single discrete capture cross section ($K=1$), the dotted line in Fig. 11 shows that it does not agree with the data. The trapping occurs over a much larger range of N_{ic} than that covered by a single capture cross section. Figure 11 shows that better agreement can be obtained by using either two discrete capture cross sections ($K=2$) in Eq. (16) or a distributed capture cross section model, Eq. (17). In both cases, the number of adjustable parameters increased when compared with a single discrete capture cross

section. The question is whether the improved agreement is a consequence of this increase of adjustable parameters. It is true that a good agreement with data can always be obtained if a large number of adjustable parameters are used and this is the reason why many researchers have their reservations with these models. In the following, we attempt to show that the improved agreement is genuine and to find out if the capture cross section is distributed or discrete.

The two discrete capture cross sections extracted are in the order of $\sigma_1 = 10^{-14}$ cm^2 and $\sigma_2 = 10^{-16}$ cm^2, respectively, so that they are well separated. If they indeed exist, it is likely that they originate from two different types of as-grown electron traps. If this is true, the two types of traps may behave differently when the test condition changes. This is indeed observed in Figs. 12(a) and 12(b). Figure 12(a) shows that a reduction of trap filling time t_{top} results in a progressive reduction of trapping in the relatively low range of N_{ic} (10^{12}–10^{15} cm^{-2}). However, further trapping in the relatively high range of N_{ic} ($>10^{15}$ cm^{-2}) is hardly affected. To show this clearly, Fig. 12(b) plots the extracted two saturation trapping levels, ΔV_{g1} and ΔV_{g2}, corresponding to the capture cross section σ_1 and σ_2, given in Fig. 12(c), respectively.

ΔV_{g1} decreases for $t_{top} < 10$ μs and becomes negligible when $t_{top} = 1.4$ μs, but ΔV_{g2} changes little over this time range. As a result, the transition between ΔV_{g1} and ΔV_{g2} is "discontinuous." It is difficult to explain this discontinuous behavior based on a continuous distribution of capture cross sections. The result strongly supports the existence of two types of traps with two discrete capture cross sections. For a 4 nm ALD HfO$_2$ layer, the effective density is around 5.3 $\times 10^{12}$ and 1.1×10^{13} cm^{-2} for the traps with a capture cross section in the order of 10^{-14} and 10^{-16} cm^2, respectively.
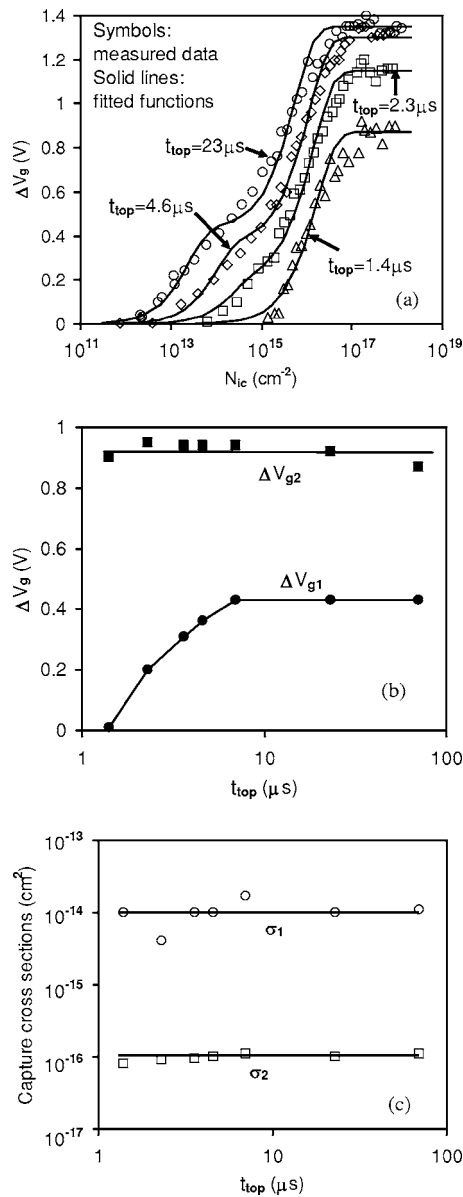
FIG. 12. Dependence of trapping on the filling time, t_{top}. (a) shows the trapping can be reduced with the filling time. The solid curves were obtained by fitting data with two discrete capture cross sections. (b) and (c) show the extracted saturation levels and capture cross sections, respectively. The extracted two capture cross sections are in the order of 10^{-14} and 10^{-16} cm^2, respectively, and they are insensitive to t_{top}. In (b), the trapping level by the smaller trap, ΔV_{g2}, which controls further trapping for $N_{ic} > 10^{15}$ cm^{-2} in (a), does not depend on t_{top} either. In contrast, the trapping level by the larger trap, ΔV_{g1}, which corresponds to the region of $N_{ic} < 10^{15}$ cm^{-2} in (a), clearly reduces for smaller t_{top}.

C. Dependence on fabrication processes and techniques

After extracting the effective density and capture cross section for electron traps, we now explore the sensitivity of these trap properties to the fabrication processes and techniques. Before comparing results reported for different samples, it should be emphasized that, even for the same test sample, both the apparent effective density and capture cross section are highly sensitive to measurement conditions, such as the measurement time, gate bias, and trap filling time. For example, Fig. 4 shows that an increase of measurement time
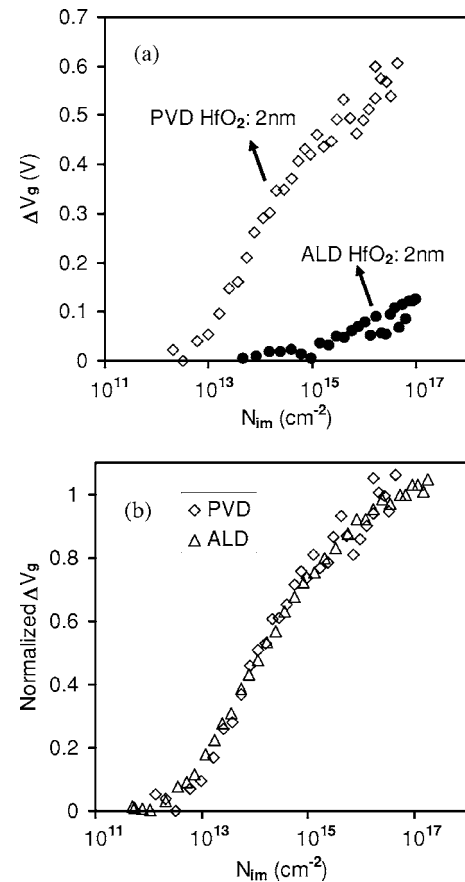

FIG. 13. A comparison of electron trapping in HfO$_2$ prepared by ALD and PVD. (a) shows that trapping in a 2 nm PVD HfO$_2$ is substantially higher than that in a 2 nm ALD HfO$_2$. (b) shows that the trapping kinetics in a 2 nm PVD HfO$_2$ is similar to that in a 4 nm ALD HfO$_2$. The trapping for ALD sample is taken from a 4 nm layer, since the trapping in a 2 nm ALD sample is too low to give reliable trapping kinetics.

can reduce both the apparent density and capture cross section significantly. Thus, any comparison with results reported by other groups must be made under the same measurement conditions. It is also reported that trapping decreases for thinner HfO$_2$ layer.[27,28] This thickness effect must be taken into account when making comparison.

It is found that the effective density obtained in our samples is within a factor of 2 of the values reported by earlier works for HfO$_2$ prepared by ALD (Refs. 29 and 30) and metal organic chemical vapor deposition (MOCVD).[5,31] For other fabrication techniques, the differences in trap density can be larger. For example, Fig. 13(a) shows that the trapping in a 2 nm HfO$_2$ prepared by PVD technique can be substantially higher than the trapping in a 2 nm ALD HfO$_2$. At this stage, we consider that the effective trap densities reported here are typical values only for ALD and MOCVD HfO$_2$ layers.

On the capture cross section, we can reproduce the values in the range of $10^{-12}-10^{-19}$ cm^2 reported earlier[3,5] on our samples if the same measurement and formula are used. The capture cross section is similar for ALD (Ref. 3) and MOCVD (Ref. 5) samples. Furthermore, Fig. 13(b) compares the trapping kinetics in PVD and ALD layers. It is obvious that there is little difference in these two. This suggests that, unlike the trap density, the capture cross section is

insensitive to fabrication processes and techniques. The wide range of the apparent capture cross sections, 10^{-12}–10^{-19} cm^2, reported in early works[3,5] mainly originates from the difference in measurement techniques and conditions as shown in Fig. 4 rather than the difference in fabrication processes and techniques.

V. CONCLUSIONS

In this work, the capture cross sections of as-grown electron traps in HfO$_2$ layer are determined. The appropriate test conditions have been chosen to suppress detrapping. The range of electron fluency has been selected so that trapping is negligible at the start, but saturates at the end. An important part of this work is estimating the trapping-induced transient gate current when a pulse is applied to the gate. This is achieved by numerical simulation. It is found that the trapping can reduce the gate current by two orders of magnitude and the gate current can drop significantly within microseconds. Without taking this transient of gate current into account, the electron fluency can be underestimated by over one order of magnitude.

To extract capture cross sections, trapping models with discrete and distributed capture cross sections are considered. It is found that trapping with the larger capture cross section decreases for a shorter filling time, but trapping with the smaller capture cross section is hardly affected over the same time range. This discontinuous behavior is against the model based on continuously distributed capture cross sections. The results support the trapping model with two well separated discrete capture cross sections, indicating the presence of two different types of as-grown electron traps. The extracted values are in the order of 10^{-14} and 10^{-16} cm^2, respectively. For a 4 nm ALD HfO$_2$ layer, the effective trap densities are around 5.3×10^{12} and 1.1×10^{13} cm^{-2} for the larger and smaller traps, respectively. These densities are typical values for 4 nm HfO$_2$ layers prepared by ALD and MOCVD techniques, but trap density can vary substantially for HfO$_2$ fabricated by other techniques, such as PVD. In contrast, it is found that the capture cross section is insensitive to the fabrication technique.

ACKNOWLEDGMENT

APPENDIX: CALCULATION OF ELECTRON FLUENCY

In this Appendix, Eq. (1) is derived for the calculation of electron fluency injected from the substrate into the dielectric, N_{im}. Since Eq. (1) is applicable for both a single layer and a stack of dielectrics, to simplify the derivation, the dielectric stack is considered as a single layer of SiO$_2$ with a thickness of EOT. As illustrated in Fig. 14, if we define $f(x)$ as the number of electrons passing through a unit area per unit time at a distance of x from the substrate/dielectric interface, the electron conservation equation is
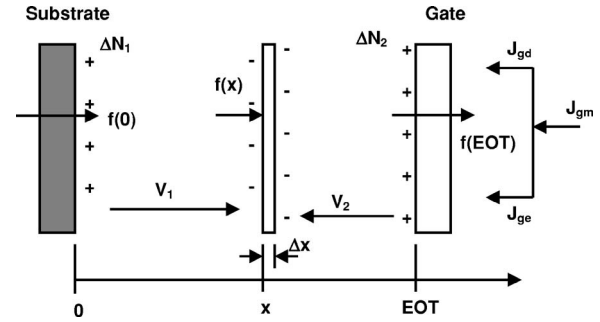

FIG. 14. A schematic illustration of electron injection, trapping, and the relevant current components used for calculating electron fluency in the Appendix.

$$\frac{\partial f}{\partial x} = -\frac{\partial \rho}{\partial t},$$

where ρ is the number of trapped electrons per unit volume. An integration against x leads to

$$f(0) = f(\text{EOT}) + \int_0^{\text{EOT}} \frac{\partial \rho}{\partial t} dx. \tag{A1}$$

The second term on the right hand side represents the loss of electrons through trapping in dielectric. By definition, N_{im} is related to $f(0)$ through

$$N_{im} = \int_0^t f(0) dt. \tag{A2}$$

The current density caused by electron flowing through the gate/dielectric interface, J_{ge}, is related to $f(\text{EOT})$ by

$$J_{ge} = q f(\text{EOT}). \tag{A3}$$

However, J_{ge} is not the measured gate current density J_{gm}, since displacement current density, J_{gd}, caused by charge trapping, also contributes to J_{gm},

$$J_{gm} = J_{ge} + J_{gd}. \tag{A4}$$

As a result, to obtain J_{ge} from the measured J_{gm}, one must evaluate J_{gd}.

As shown in Fig. 14, within a thin layer of dielectric, Δx, the trapped electrons per unit area are

$$\Delta N = \rho \Delta x.$$

The ΔN induces a positive charge of ΔN_1 at the substrate interface and ΔN_2 at the gate interface with

$$\Delta N = \Delta N_1 + \Delta N_2. \tag{A5}$$

The corresponding potential drop between the substrate and the layer at x is

$$V_1 = \frac{q \Delta N_1 x}{\varepsilon_0 \varepsilon_{\text{SiO}_2}}, \tag{A6}$$

and similarly,

$$V_2 = \frac{q\Delta N_2(\text{EOT} - x)}{\varepsilon_0 \varepsilon_{\text{SiO}_2}}. \tag{A7}$$

Since the test was carried out with a constant voltage over the dielectric, it requires

$$V_1 = V_2. \tag{A8}$$

Solving Eqs. (A5)–(A8), we have

$$\Delta N_2 = \frac{x}{\text{EOT}}\Delta N = \frac{x}{\text{EOT}}\rho\Delta x. \tag{A9}$$

Integrating Eq. (A9), the total trapping-induced positive charges on the gate is

$$N_g = \int_0^{\text{EOT}} \frac{x}{\text{EOT}}\rho dx.$$

The displacement current is

$$J_{\text{gd}} = q\frac{\partial N_g}{\partial t} = q\frac{\partial}{\partial t}\left(\int_0^{\text{EOT}} \frac{x}{\text{EOT}}\rho dx\right). \tag{A10}$$

By using Eqs. (A3), (A4), and (A10), the $f(\text{EOT})$ can now be evaluated through

$$f(\text{EOT}) = \frac{1}{q}(J_{\text{gm}} - J_{\text{gd}}) = \frac{J_{\text{gm}}}{q} - \frac{\partial}{\partial t}\left(\int_0^{\text{EOT}} \frac{x}{\text{EOT}}\rho dx\right).$$

From Eq. (A1), we have

$$\begin{aligned}
f(0) &= f(\text{EOT}) + \int_0^{\text{EOT}} \frac{\partial\rho}{\partial t}dx \\
&= \frac{J_{\text{gm}}}{q} - \frac{\partial}{\partial t}\left(\int_0^{\text{EOT}} \frac{x}{\text{EOT}}\rho dx\right) + \int_0^{\text{EOT}} \frac{\partial\rho}{\partial t}dx \\
&= \frac{J_{\text{gm}}}{q} + \frac{\partial}{\partial t}\left[\int_0^{\text{EOT}}\left(1 - \frac{x}{\text{EOT}}\right)\rho dx\right].
\end{aligned}$$

Using Eq. (A2), the electron fluency at the substrate/dielectric interface is

$$N_{\text{im}} = \int_0^t f(0)dt = \frac{1}{q}\int_0^t J_{\text{gm}}dt + \int_0^{\text{EOT}}\left(1 - \frac{x}{\text{EOT}}\right)\rho dx. \tag{A11}$$

For a fixed drain current, the trapping induced gate voltage shift is

$$\Delta V_g = \frac{q}{C_{\text{ox}}}\int_0^{\text{EOT}}\left(1 - \frac{x}{\text{EOT}}\right)\rho dx. \tag{A12}$$

Combining Eqs. (A11) and (A12), we have

$$N_{\text{im}} = \frac{1}{q}\int_0^t J_{\text{gm}}dt + \frac{\Delta V_g C_{\text{ox}}}{q}. \tag{A13}$$

If one assumes J_{gm} is a constant over a period of t_{top}, Eq. (A13) becomes our Eq. (1),

$$N_{\text{im}} = J_{\text{gm}} \times t_{\text{top}}/q + \Delta V_g \times C_{\text{ox}}/q.$$

[1] G. Groeseneken *et al.*, in Proceedings of the 11th International Symposium on Physical and Failure Analysis of IC, Hsinchu, Taiwan, 2004 (IEEE, New York, 2004), pp. 147–155.

[2] T. P. Ma, H. M. Bu, X. W. Wang, L. Y. Song, W. He, M. Wang, H. -H. Tseng, and P. J. Tobin, IEEE Trans. Device Mater. Reliab. **5**, 36 (2005).

[3] S. Zafar, A. Callegari, E. Gusev, and M. V. Fischetti, J. Appl. Phys. **93**, 9298 (2003).

[4] C. Z. Zhao, M. B. Zahid, J. F. Zhang, G. Groeseneken, R. Degraeve, and S. De Gendt, Microelectron. Eng. **80**, 366 (2005).

[5] G. Bersuker, J. H. Sim, C. D. Young, R. Choi, P. M. Zeitzoff, G. A. Brown, B. H. Lee, and R. W. Murto, Microelectron. Reliab. **44**, 1509 (2004).

[6] R. Degraeve, A. Kerber, Ph. Roussel, E. Cartier, T. Kauerauf, L. Pantisano, and G. Groeseneken, Tech. Dig. - Int. Electron Devices Meet. **2003**, 935.

[7] Z. Xu, M. Houssa, S. De Gendt, and M. Heyns, Appl. Phys. Lett. **80**, 1975 (2002).

[8] A. Kerber *et al.*, Proceedings of IEEE 41st Annual Int. Reliab. Phys. Symp., Dallas, Texas, 2003, pp. 41–45.

[9] A. Y. Kang, P. M. Lenahan, and J. F. Conley, Appl. Phys. Lett. **83**, 3407 (2003).

[10] G. Reimbold, J. Mitard, M. Casse, X. Carros, C. Leroux, L. Thevenod, and F. Martin, in Proceedings of the Silicon Nitride, Silicon Dioxide Thin Insulating Films, and Other Emerging Dielectrics VIII, Quebec City, Canada, 2005 (Electrochemical Society, NJ, 2005), pp. 437–455.

[11] J. F. Zhang, S. Taylor, and W. Eccleston, J. Appl. Phys. **71**, 725 (1992).

[12] T. Schram *et al.*, Microelectron. Reliab. **45**, 779 (2005).

[13] W. D. Zhang, J. F. Zhang, M. Lalor, D. Burton, G. Groeseneken, and R. Degraeve, IEEE Trans. Electron Devices **49**, 1868 (2002).

[14] J. F. Zhang, S. Taylor, and W. Eccleston, J. Appl. Phys. **71**, 5989 (1992).

[15] M. Houssa, A. Stesmans, and M. M. Heyns, Semicond. Sci. Technol. **16**, 427 (2001).

[16] Y. Zhao and M. H. White, Solid-State Electron. **48**, 1801 (2004).

[17] B. Govoreanu, P. Blomme, K. Henson, J. V. Houdt, and K. D. Meyer, Solid-State Electron. **48**, 617 (2004).

[18] N. Yang, W. K. Henson, J. R. Hauser, and J. J. Wortman, IEEE Trans. Electron Devices **46**, 1464 (1999).

[19] K. N. Yang *et al.*, IEEE Trans. Electron Devices **47**, 2161 (2000).

[20] S. Mudanai, Y. Y. Fan, Q. Ouyang, A. F. Tasch, and S. K. Banerjee, IEEE Trans. Electron Devices **47**, 1851 (2000).

[21] Y. T. Hou, M. F. Li, H. Y. Yu, and D. L. Kwong, IEEE Electron Device Lett. **24**, 96 (2003).

[22] E. M. Vogel, K. Z. Ahmed, B. Hornung, W. K. Henson, P. K. McLarty, G. Lucovsky, J. R. Hauser, and J. J. Wortman, IEEE Trans. Electron Devices **45**, 1350 (1998).

[23] M. L. Roy, E. Lheurette, O. Vanbesien, and D. Lippens, J. Appl. Phys. **93**, 2966 (2003).

[24] Y. Y. Fan, R. E. Nieh, J. C. Lee, G. Lucovsky, G. A. Brown, L. F. Register, and S. K. Banerjee, IEEE Trans. Electron Devices **49**, 1969 (2002).

[25] M. H. Chang, J. F. Zhang, and W. D. Zhang, IEEE Trans. Electron Devices **53**, 1347 (2006).

[26] A. Shanware *et al.*, Tech. Dig. - Int. Electron Devices Meet. **2003**, 939.

[27] J. H. Sim, S. C. Song, P. D. Kirsch, C. D. Young, R. Choi, D. L. Kwong, B. H. Lee, and G. Bersuker, Microelectron. Eng. **80**, 218 (2005).

[28] J. F. Zhang, C. Z. Zhao, M. B. Zahid, G. Groseneken, R. Degraeve, and S. De Gendt, IEEE Electron Device Lett. **27**, 817 (2006).

[29] C. D. Young *et al.*, Microelectron. Reliab. **45**, 806 (2005).

[30] C. Leroux, J. Mitard, G. Ghibaudo, X. Garros, G. Reimbold, B. Guillaumot, and F. Martin, Tech. Dig. - Int. Electron Devices Meet. **2004**, 737.

[31] C. Shen *et al.*, Tech. Dig. - Int. Electron Devices Meet. **2004**, 733.