Extraction of Si-SiO\textsubscript{2} Interface Trap Densities in MOSFET’s With Oxides Down to 1.3 nm Thick

By D. Bauza, Institut de Microelectronique, Electromagnétisme et Photonique (IMEP), UMR CNRS 5130, INPG, ENSERG, 23, rue des Martyrs, BP 257, 38016 Grenoble Cedex 1, FRANCE.
E-mail: bauza@enserg.fr

Abstract

It is shown that reliable interface trap density values can be extracted from MOS devices with ultrathin oxides by using charge pumping and small gate voltage swings. This presents three advantages with respect to the conventional large gate voltage swing approach: the extraction is simple as carrier emission do not contribute to the CP signal so that the CP current magnitude directly reflects the interface trap density; the tunneling current is strongly reduced allowing a more easy extraction of the CP signal; this prevents the insulator and the insulator-silicon interface from any degradation. By doing so, interface trap densities from MOS devices with oxides down to 1.3 nm thick are reported for the first time.

1. Introduction

In VLSI MOS devices, conventional interface trap characterizations techniques become unusable as soon as the parasitic Fowler-Nordheim or direct tunneling currents become significant. This is an important feature, as interface trap density, $D_\text{it}$, measurements have always been a standard characterization tool for technological processes. For this, the use of charge pumping (CP) at high frequencies has been proposed but the ratio CP over leakage current remains small [1]. Methods based on stress induced leakage current are also studied [2, 3]. In this article, $D_\text{it}$ is measured simply with CP using small gate voltage pulses. This strongly increases the frequency range in which the CP signal can be measured and allows reliable $D_\text{it}$ values to be obtained.

2. Extraction of interface trap densities using small gate voltage pulses

Wachnik and Lowney [4] have proposed a model of CP in which small gate pulses with abrupt transition edges are assumed. This model holds at the maximum of Elliott curves [5], in which the gate bias is scanned while keeping the gate pulse amplitude constant. It is based on the fact that the traps located at Fermi level during the time at the high and low bias levels have identical emission and capture rates so that when switching the device from one level to the other, the traps that could emit their carrier towards the nearest carrier band, capture a carrier from the other carrier band and emission do not contribute to the CP current.

Accounting for the large distribution of trap time constants at the Si-SiO\textsubscript{2} interface, the charge recombining during one period of the gate signal, $Q_\text{cp}$, can be written as [6, 7]:

$$Q_\text{cp} = \frac{I_\text{cp}}{f} = qWL \int_{E_\text{hi}}^{E_\text{lo}} \int_{0}^{d_{\text{ox}}} N_t(E,x) \Delta F(E,x) dE$$

(1)

where $I_\text{cp}$ is the CP current, $f$ is the gate signal frequency, $q$ is the absolute electron charge, $W$ and $L$ are the width and length of the device, $E_\text{hi}$ and $E_\text{lo}$ are the Fermi level position at the interface during the high and low bias, $d_{\text{ox}}$ is the gate oxide thickness, $N_t(E,x)$ is the volumic trap concentration at energy $E$ and distance $x$ from the Si-SiO\textsubscript{2} interface when assuming tunneling for the capture of carriers [6, 7]. $\Delta F(E)$, the filling function variation of the traps primarily derived by Wachnik and Lowney [4], has been modified to account for the large trap time constant distribution [6, 7]:

$$\Delta F(E,x) = \frac{\left[1 - \exp\left(- \frac{c_{n,p}(E,x) t_{h,l}}{1 - \exp\left(- 2c_{n,p}(E,x) t_{c,n,p}\right)}\right)^2\right]}{1 - \exp\left(- 2c_{n,p}(E,x) t_{c,n,p}\right)}$$

(2)

In equation 2, $t_{h,l} = (1/2f - 2 t_{c,n,p})$ is the time for capture, $t_{c,n,p} = t_{c} = t_{l}$, being the transition time of the gate signal. $c_{n,p}(E,x)$ is the capture rate for electrons and
holes of the traps at E and x. \( c_n(E, x) = c_p(E, x) \) are equal at Elliot maxima \[7\]:

\[
c_{n,p}(E, x) = c_n(E, x) = c_p(E, x) = (n=p) \sigma_{c,h}(x) v_{th} \quad (3),
\]

with:

\[
\sigma_{c,h}(x) = \sigma_{c,h}(0) \exp(-x/\lambda) \quad (4).
\]

\( n \) and \( p \) are the electron and hole concentration at the silicon surface, \( \sigma_{c,h}(x) = \sigma_c(x) = \sigma_h(x) \) is the trap capture cross sections for electrons and holes at \( x \), \( v_{th} \) is the average carrier thermal velocity, and \( \lambda \) is the tunnel attenuation distance \[8\].

In practical situations, the transition times of the gate signal cannot be overlooked. Emission does not contribute to the CP current if the trap that could emit a carrier after the whole transition time are situated outside the interval \( (E_h - E_l) \).

From Shockley-Read-Hall kinetics, this means that:

\[
E_{em,e} - E_i = -kT \ln(n_i \sigma_e v_{th} t_f) > E_h - E_i \quad (5),
\]

and

\[
E_i - E_{em,h} = -kT \ln(n_i \sigma_h v_{th} t_r) > E_i - E_i \quad (6),
\]

where \( E_{em,e} \) and \( E_{em,h} \) are the energy position of the traps that can emit after \( t_f \) and \( t_r \), \( E_i \) is the intrinsic level, \( n_i \) is the intrinsic carrier density, \( k \) and \( T \) being the Boltzmann constant and the temperature, respectively.

As a result, when \( t_f, t_r = 50 \text{ ns} \) in figure 2, as used in the experiments, equation 1 holds up to \( V_{sw} = 0.7 \text{ to } 1 \text{ V} \) depending on the trap cross sections and provided that most of the traps are filled, that is, \( \Delta F(E,x) = 1 \) in equation 1 as far as possible towards the oxide depth \[7\].

In the small gate pulse approach, saturation of the traps with carriers and therefore saturation of \( Q_{cp} \) can be obtained by increasing the time for capture, that is, reducing the gate signal frequency.
Then, the areal interface trap density, \( D_{it} \), can be very simply obtained from equation 1 and from the CP signal magnitude as:

\[
D_{it} = \frac{Q_{cp}}{qWL(E_h - E_f)} \quad (7)
\]

3. Results

The devices used are n-channel transistors fabricated at CPMA-Grenoble, France. Their width and length were 300 x 0.2 \( \mu \)m\(^2\) for the devices with \( d_{ox} = 2.3 \) and 1.3 nm and 10 x 10 \( \mu \)m\(^2\) for the devices with \( d_{ox} = 1.8 \) nm. For these later devices, the transition time of the gate signal was increased from 50 ns to 300 ns to suppress the geometric effect [9]. The doping concentrations were \( 3 \times 10^{17} \) cm\(^{-3}\), \( 6 \times 10^{17} \) cm\(^{-3}\) and \( 2 \times 10^{18} \) cm\(^{-3}\) for the devices with \( d_{ox} = 2.3, 1.8 \) and 1.3 nm, respectively.

Figure 3 shows Elliot curves recorded at 1000 Hz for \( V_{sw} \) values comprised between 0.3 and 1 V from a device with \( d_{ox} = 1.8 \) nm. With such small gate pulses, the CP current can be easily observed down to such frequencies. Nevertheless, the leakage current must be removed by subtracting CP curves recorded at 50 Hz in which the CP signal is negligible [1].

Figure 4 shows the evolution of \( D_{it} \) with \( f \) and \( V_{sw} \) for different \( d_{ox} \) when \( d_{ox} = 2.3 \) and 1.8 nm is displayed. As expected, \( D_{it} \) saturates when increasing \( V_{sw} \) and reducing \( f \) so that from the CP curves at 0.8 or 1 V, \( D_{it} = (1.4 \times 10^{10}) \) eV\(^{-1}\) cm\(^{-2}\) and (1.8-2 \( \times 10^{10} \)) eV\(^{-1}\) cm\(^{-2}\) for these technologies.

The slight decrease of \( D_{it} \) for the largest \( V_{sw} \)'s at low frequencies results from the small CP current that exists in the curves at 50 Hz used to suppress the leakage component or by the inaccuracy of this method for a very precise extraction of the CP current.

Figs. 5 and 6 show the results obtained for a device with \( d_{ox} = 1.3 \) nm. The CP current is now swamped in the leakage current (Fig. 5). Nevertheless, Elliot curves can be extracted at high frequency, as also shown in Fig 5, so that for this technology, \( D_{it} = (2 \times 10^{12}) \) eV\(^{-1}\) cm\(^{-2}\) as shown in Fig. 6.

One can note in Fig. 6, that the dependence of \( D_{it} \) with \( V_{sw} \) is weaker for this oxide thickness than for the thicker oxides in Fig. 4. This suggests large trap capture cross sections for this oxide. In addition, the increase of \( D_{it} \) between \( d_{ox} = 1.8 \) and 1.3 nm is much greater than between \( d_{ox} = 2.3 \) and 1.8 nm. This could result from the
increase of the stress at the Si-SiO\textsubscript{2} interface when reducing $d_{ox}$, in relation with the reduction of the time to breakdown [10].

![Graph]

Figure 6. $D_{it}$ obtained as a function of $f$ and $V_{sw}$ for a device with $d_{ox} = 1.3$ nm

### 4. Conclusion

It has been shown that reliable interface trap densities values can be measured in MOS devices with ultrathin oxides. This is achieved by using charges pumping in the small gate pulse mode and has been applied to devices with different oxide thickness, down to 1.3 nm. This prevents the oxides from any degradations and strongly reduces the oxide leakage allowing the extraction of the CP current in a large range of gate signal frequencies and/or down to very small thicknesses.

### 5. References


This work has been supported by Ultimox/RMNT and RA-Nanodiel projects. The author wish to thank Martine Gri from IMEP for technical assistance.