Correlation between Oxide Breakdown Field and Effective Oxide Charge Concentration in Ultra-Thin Tunnel Oxide of FG Flash Memory

R.M. Ayub, S. Norhafizah, A.H. Azman, U. Hashim, Tijjani Adam
Institute of Nano Electronic Engineering, Universiti Malaysia Perlis (UniMAP), 01000 Kangar, Perlis Malaysia.

ABSTRACT

Experiments were carried out to study the relationship between the oxide breakdown field of thin silicon dioxide (SiO₂) film, and the effective oxide charge concentration, \( N_{\text{EFP}} \). The Metal-Oxide-Semiconductor (MOS) capacitors were used as the test vehicle where standard MOS fabrication process technologies were employed. The main device fabrication processes involved were dry thermal oxidation, photolithography, etching and metallization. After the final sintering process, the electrical device characterization namely the Current-Voltage (I-V) and Capacitance-Voltage (C-V) was carried out on MOS capacitor test structures. The result shows that the oxide's breakdown field is significantly correlated with the amount of the effective oxide charge for the whole studied thickness range. The breakdown field also exhibited power-law dependency on the oxide thickness. Therefore, it can be qualitatively concluded that the amount of trapped charge in SiO₂ would promote the trap-assisted tunneling component and subsequently, causing an oxide breakdown.

Key words: Oxide breakdown, tunnel oxide, SILC, effective oxide charge, floating gate flash memory.

Introduction

Floating gate device is the mainstream of flash memory and mostly used in both NOR and NAND cell architectures. The ubiquitous presence of NAND in almost all aspect of modern life especially, has led the flash memory device scaling towards 10 nm technology node with blistering speed, surpassing both logic and DRAM (Lu, C.Y., 2012). In this device, the charge is stored on a conducting layer that is completely isolated from other structures by a very thin dielectric film. The operational of floating gate device is based upon the ability to bring electrons onto the floating gate and removing them again in order to change the threshold voltage of the memory cell. The pace at which these operations can be carried out is the most important floating gate device performance indicator and its normally termed as the programming speed (Wellekens, D. and J. Van Houdt, 2008; Wellekens, D. and J. Van Houdt, 2008).

The electron injection is carried out via ultra-thin dielectric layer, called the tunnel oxide, which transport the electrons under the influence of external electric field. Generally, the higher the electric field across the tunnel oxide, the higher the rate of electron injection through it. Since the invention of flash memory, the thickness of tunnel oxide has been scaled down aggressively to improve the programming speed. However, for the tunnel oxide thickness of less than 8 nm, numbers of unwanted phenomena, most importantly the stress-induced leakage current (SILC) has been reported (Wellekens, D. and J. Van Houdt, 2008; Shin, Y., 2010; Lu, C.Y., 2009). The SILC has been shown to severely affecting another important aspect of floating flash performance, namely the data retention time. The 2011 ITRS report shows that the working tunnel oxide thickness ranging from 8 down to 4 nm is very critical for NAND’s technology nodes of 28 nm towards 16 nm (ITRS, 2011). Furthermore, it has been reported that the oxide early breakdown is highly correlated with the generation of trap at Si/SiO₂ interface under high field stress, which in turn would promote the low field leakage, which is the SILC (Connor, R.O., 2011; Yamada, R., 2000; Yang, J., 2006). Therefore, this study is carried out to investigate the relationship between the oxide breakdown and the amount of effective oxide trap charge.

Experimental Details:

Device Fabrication:

Conventional aluminum gate MOS capacitors with blanket tunnel oxides were employed as the test structure. The 100 mm diameter, <100>, p-type silicon wafers with the resistivities of 1-20 Ohm-cm were used.
as the starting wafers. The wafers were processed through standard wafer cleaning such as RCA1 and RCA2 to remove metallic and organic contaminants. HF dip is done for approximately 10 s to remove the native oxides. The tunnel oxides were thermally grown in a dry O2 ambient at a temperature of 850°C, with thickness ranging from 2 to 12 nm as measured by Filmetrics F20-UV, and later verified by the C-V measurement. Aluminum (approximately 0.4 μm) was deposited at the front and the back of the wafer by thermal evaporation technique. The gate metal is then patterned with the chrome-type mask using the MDA-400M mask aligner system. The wafers were then etched and stripped to remove the unwanted materials. Finally, the wafers were sintered in N2 ambient at 400°C for 30 minutes. 

**I-V Characterization** - The Keithley 4200-SCS Parameter Analyzer is used to perform the voltage ramp dielectric breakdown (V-RAMP) test. The oxide breakdown voltages are determined from the change in the slope of tunnel oxide current versus gate voltage, where the value of 3 is used as the exit value as recommended by JEDEC document (Standard, J., 1992).

**C-V Characterization** - Alpha-A-High Performance Frequency Analyzer was used to perform the high frequency CV analysis (100 MHz). From the CV curves, the effective oxide charge is calculated using procedures outlined in Nicollian and Brews (Nicollian, E.H. and J.R. Brews, 1982), with the assumption that the charge is located in a sheet at the silicon-to-silicon dioxide interface. 

Based on this procedure, firstly the flatband voltage is extracted from the CV curve by the flatband capacitance method, $C_{FB}$ based on the relationship:

$$C_{FB} = \frac{C_{ox} \cdot \epsilon_0 \cdot A}{\epsilon_0 \cdot \epsilon_0 \cdot A / \lambda_D}$$  \hspace{1cm} (1)

Where $C_{FB}$ is the flatband capacitance (pF), $C_{ox}$ is the tunnel oxide capacitance, $\epsilon_0$ and $\epsilon_{Si}$ is the vacuum and substrate material permittivity respectively (F/cm), A is the gate area (cm²), and $\lambda_D$ is the extrinsic Debye length, which can be calculated from:

$$\lambda_D = \left[ \frac{q^2 \cdot \epsilon_{Si} \cdot kT}{\epsilon_0 \cdot q^2 \cdot N_x} \right]^{1/2}$$  \hspace{1cm} (2)

Where $kT$ is the thermal energy at room temperature (4.046 x 10^{-21}J), q is the electron charge (1.6 x 10^{-19} C), and $N_x = N$ at 90% $W_{MAX}$, or $N_x = N_{i} = N_{D}$. Once the value of $C_{FB}$ is known, the value of $V_{FB}$ can be obtained from the C-V curve data as shown in Figure 3. The effective oxide charge, $Q_{EFF}$ can be calculated from;

$$V_{FB} - W_{MS} = - \frac{Q_{EFF}}{C_{ox}}$$  \hspace{1cm} (3)

The $W_{MS}$ value is assumed to be -0.95 V for this structure (Mos-capacitor, U.). Finally, $N_{EFF}$ can be calculated from;

$$N_{EFF} = \frac{Q_{EFF}}{q}$$  \hspace{1cm} (4)

**Results:**

**Tunnel Oxide Breakdown:**

Typical I-V curves for MOS capacitors with various tunnel oxide thickness stressed under inversion condition is shown in Figure 1. The V_RAMP tests were performed on those capacitors until oxides breakdown were achieved. The breakdown fields were then calculated and plotted against tunnel oxide thickness as in Figure 2.

![Fig. 1: Gate Current versus Gate Voltage.](image-url)
The gate leakage current is composed of electron tunneling from the conduction band (ECB), electron tunneling from the valence band (EVB) and hole tunneling from the valence band (HVB) (Lee, W. and C. Hu, 2001). Furthermore, it has been widely reported that for oxide below 4 nm, the direct tunneling through the oxide’s trapezoidal energy barrier would become dominant. It’s also shown that the existence of oxide trap would promote another component of tunneling namely trap-assisted tunneling (Ghetti, A., 2001). Qualitatively, we can explain the behaviour of I-V curves in Figure 1 as follow;

- **For 2, 3, 4 and 6 nm oxides**: besides the tunneling current contributed by the F-N tunneling, the other dominant tunneling mechanism is direct tunneling. These would explain two distinct regions of curves in Figure 1.
- **For 8 and 12 nm oxides**: the main tunneling mechanism is the F-N tunneling with some portion of hot electron injection.
- **For 2 nm oxide**: due to high concentration of $N_{eff}$ as shown in Figure 4, additional tunneling mechanism is anticipated in the form of trap-assisted tunneling. This would explain the observed high leakage current.
- To quantitatively separating the tunneling mechanisms involved, a detailed quantum mechanical modeling of the whole device structure is required (Ghetti, A., 2001).

The dotted line in Figure 2 is the trend-line calculated using MS Excel to show the power-law dependence of the oxide breakdown field on oxide thickness with acceptable $R^2$ value. This observed trend fits well with the power-law model that has been successfully adopted to explain the ultra-thin oxide breakdown phenomenon (Ohgata, K. and M. Ogasawara, 2005).

![Fig. 2: Breakdown Field versus Tunnel Oxide Thickness.](image)

**Effective Oxide Charge Concentration:**

Charge trapping is an important issue in oxide reliability since high charge trapping in oxide eventually leads to oxide breakdown. The effective oxide charge $Q_{eff}$ represents the sum of the oxide fixed charge $Q_f$, the mobile charge $Q_m$, and the oxide trapped charge $Q_{ot}$. These three components can be distinguished from each other by temperature cycling test. $Q_{eff}$ is distinguished from the interface trapped charge $Q_{it}$ in that $Q_{it}$ varies with gate bias, whereas $Q_{eff}$ does not.

Figure 3 shows a typical high frequency C-V curves plotted for MOS capacitor (2 nm tunnel oxide in this case). The device is swept from an accumulation to the inversion region, giving the corresponding capacitance values as shown by the dotted line. $C_{ox}$ is the maximum capacitance at the accumulation region. $C_{fb}$ is then calculated from Equation 1 and then $V_{fb}$ is extracted using procedure as outlined in device characterization section as shown by the dotted line. The values of $N_{eff}$ were then calculated and plotted against tunnel oxide thickness as shown in Figure 4. A strong dependence of $N_{eff}$ on the tunnel oxide thickness has been observed. Theoretically, $N_{eff}$ will be strongly dependence on the processing parameters, especially on the oxidation temperature and post oxidation annealing process. For this experiment, all the tunnel oxides were grown and annealed at the same temperatures. Therefore, we did not anticipate any significant variation of $N_{eff}$ over the tunnel oxide thickness.
The typical $N_{\text{eff}}$ value in as grown thermal SiO$_2$ is in the high level of $10^9$ charge per cm$^3$, as shown for 8 nm and 12 nm tunnel oxide which are $5.1 \times 10^9$ and $1.4 \times 10^{10}$ charge per cm$^3$ respectively. The $N_{\text{eff}}$ values in 3, 4 and 6 nm tunnel oxides were two to five times higher compared to that of 8 and 12 nm tunnel oxides, and more than ten times higher in 2 nm tunnel oxides.

As been widely accepted, high leakage current observed in thin oxides is caused by the defects in oxide which act as charge trapping center. During the voltage stress, electrons gain energy and would eventually release it. The energy release mechanism is highly dependence on the oxide thickness (Irrera, F., 2005).

For oxides thickness > 4 nm, if the energy is high enough, it could be transferred to the oxide network and break Si-H bonds, thus creating new traps either at the interface or in the bulk. Therefore, the creation of trap is self-limited by the existing defects. In the case of oxides thickness < 4 nm on the other hand, electrons are ballistic. Therefore the trap creation mechanism cannot be self-limited. In addition, it’s widely reported that the trap creation mechanism would still occur even though the electron energy is lower than the Si-H bond breaking energy threshold (Stathis, J.H., 2001).

**Conclusion:**

The results show strong correlation between the oxide breakdown field, and effective oxide charge concentration. The dependence of $N_{\text{eff}}$ on the oxide thickness cannot be explained conventionally. Therefore, it has to be explained from the gate oxide reliability model. Based on this model, high-field stress would generate oxide trapped charge and the rate of trapped charge generated is increase as the tunnel oxide thickness decrease.

**Acknowledgment**

The author would like to thanks Ministry of Science, Technology and Innovation (MOSTI) for funding this research through the Science Fund project with the code number 03-01-15-SF0093, titled “Advanced Flash Memory Development for 32nm Technology Node and beyond.”
References


