A discharge-based multi-pulse technique (DMP) for probing electron trap energy distribution in high-k materials for Flash memory application

X.F.Zheng(1), W.D.Zhang(1)*(*w.zhang@ljmu.ac.uk), B. Govoreanu(2), J.F.Zhang(1), J. van Houdt(2)

(1)School of Engineering, Liverpool John Moores University, Byrom Street, Liverpool L3 3AF, UK
(2) TU/PT Division, IMEC, Leuven, B3001, Belgium

Abstract

Electron trap energy distribution in the bulk of high-k materials is the essential information required for accurate retention prediction and fast material and processing optimization in Flash memory application. A new discharge-based multi-pulse technique (DMP) has been developed in this work, which, for the first time, gives this distribution across the bulk of high-k dielectric layer. Electron trap energy distributions in HfO_2, Al_2O_3 and HfAlO stacks have been extracted and compared to identify the effects of material variation.

Introduction

The application of high-k materials, such as Al_2O_3 or HfAlO, is essential for downscaling Flash memory technology beyond the 30 nm generation. However, the progress has been held back by the large defect density in high-k materials, which limits the memory retention due to the excessive low-field leakage current induced by trap-assisted tunneling [1]. To estimate the low-field leakage current, electron trap energy distribution in the bulk of high-k materials is required, but still largely missing [1,2]. In this work, limitations in existing characterization techniques such as Charge Pumping (CP) and Charge Injection and Sensing (CIS) have been overcome by the newly developed discharge-based multi-pulse technique (DMP). The overall electron trap energy distributions across the high-k stacks, such as HfO_2, Al_2O_3 and HfAlO (Table 1), have been extracted and compared. The effects of material variation on the electron trap distribution have been identified.

Results and Discussions

A. Limitations of existing techniques

Various CP techniques have been used to characterize the energy/spatial distribution of electron traps in high-k materials [3-5]. However, CP can only reach the traps at around 2 nm from the substrate interface (Fig.1) [4,5], since its maximum discharge time is limited practically at 10 ms. Most traps in thicker high-k films used for Flash application cannot be probed. Moreover, CP with variable discharge bias can only probe traps at energy levels corresponding to Si band gap (Fig.2) [3]. A powerful technique, CIS, was developed to probe trap distribution at electron injection level through sweeping the injection bias [6]. However, as illustrated in the energy band diagram in Fig.3, only a limited region above the Si conduction band bottom (E_{CB}) can be probed. A new DMP technique has been developed in this work to overcome these limitations.

B. DMP technique

In order to probe electron traps across the high-k layer they first need to be filled. Applying a high positive bias on the gate can inject electrons into the conduction band of the high-k layer and fill the traps via scattering (Fig.4a). The filling time should be long enough so that saturation in trapping is achieved. Fig.4b shows the typical I_d-V_g measured by pulsed technique during the charging, and the saturation is reached within 100 sec (Fig.5). There are few new traps generated under the charging conditions here, as good agreement is observed between the 1st filling on a fresh device and the 2nd filling on the same device after a complete discharge (Fig.5).

After the trap filling is saturated, the gate bias is reduced to discharge the electron traps. Only electron traps with energy level higher than the silicon conduction band bottom can be discharged via tunneling since there are little empty states in the silicon band gap [3]. The energy band diagram of SiO_2/high-k gate stack during the discharging at a lower positive gate bias and a negative gate bias is shown in Fig. 6a&b, respectively. Switching to a lower gate bias will not cause further trap filling as the electron trapping in the bulk is saturated; instead, shallower traps in the white region in Fig.6 a&b become dischargeable. This is because at a lower positive bias, the potential dropped across the SiO_2 layer is reduced, which causes the trap energy level in the white region to increase to above the Si E_{CB} at the substrate interface, as shown in Fig. 6a. At a negative bias, as shown in Fig. 6b, the potential across the SiO_2 layer starts to increase towards the opposite direction. Traps at energy levels corresponding to $\Delta E_{IL}<0$ in the high-k layer become dischargeable as their energy levels are increased to above the Si E_{CB}. Deeper traps in the shadowed region remain charged as there are little empty states in Si band gap. By lowering $V_{discharge}$ sequentially and measuring the complete discharge between two successive $V_{discharge}$, trap density against energy level can be measured. Smaller steps in $\Delta V_{discharge}$ will result in finer energy level division.

The principle of the multi-pulse technique is illustrated in Fig.7a&b. Immediately after the trap filling is saturated, the gate bias is reduced to and then kept at a lower level $V_{discharge1}$

until the new saturation level of trapping is observed, which allows a complete discharge of the shallower traps, as shown in Fig.7a. The discharge is interrupted by a number of short pulses at various intervals to measure the discharged trap density from the I_d-V_g on the pulse edge. This discharging process is repeated for lower gate biases on the same device until the discharge bias is smaller than V_{th}, from there onwards, the direction of the short pulses is changed so that V_{th} can be measured, as shown in Fig.7b. Fig.7c shows the detailed test procedure. Note that during the discharging, electron trapping level changes significantly due to the high trapping density in the high-k layer. This causes V_{th} to shift accordingly and in turn changes the potential drop across the SiO_2 layer. In order to take this effect into account, each V_{th} measured at the discharging intervals is used as the monitor of the effective trapping level in the dielectric stack, and the discharge bias is adjusted accordingly to ensure V_{SiO2} and therefore ΔE_{IL} at the SiO_2/high-k interface are kept constant during each discharge stage.

Fig.8a shows the typical I_d-V_g during the discharge after the trap filling, and Fig.8b shows the extracted trapping level after charging and discharging at different $T_{discharge}$, respectively. Their difference, $D\Delta V_{th}=\Delta V_{th}$(end of charging)-$\Delta V_{th}(T_{discharge})$, is the discharge-induced V_{th} shift, which can be converted into the equivalent trap sheet density at substrate interface, ΔN_{DMP}. Given a long enough $T_{discharge}$ and a large enough negative bias, all traps throughout the stack can be discharged as $D\Delta V_{th}$ approaches the trapping level. This confirms that, unlike the CP and CIS technique, the new DMP technique allows probing traps throughout the stack. Fig.9 shows that using 5 μs for the rising/falling edge does not cause additional discharge during the measurement. The maximum negative discharge bias used here is -5 V, which does not introduce any additional trapping, as no charge trapping is observed when the negative bias of -5 V is applied on a fresh device.

C. Electron trap energy distributions

Energy distribution of electron traps across the high-k layer has been obtained by using the above technique to measure the discharged trap density against the corresponding ΔE_{IL}. The results of discharging under all given $V_{discharge}$ against $T_{discharge}$ are shown in Fig.10a for a 2nm/10nm SiON/Al_2O_3 stack. It is clear that the amount of detrapping increases when the positive discharge bias is reduced, caused by the discharge of shallow traps at the lower positive bias. The cumulative ΔN_{DMP} measured at the saturation point of each discharge against the corresponding ΔE_{IL} is shown in Fig. 10b. Nearly 60% of the traps in Al_2O_3 are below Si E_{CB}, in agreement with ref. [7]. The ΔN_{DMP} for each discharge and the corresponding energy distribution is shown in Fig.11 a&b. The small peak at 0.5 eV above Si E_{CB} agrees well with that extracted by CIS [6]. The large peak near Si E_{CB} and the deeper peak below E_{CB}, which could not be extracted by CIS, actually dominate the trapping. Since many attribute the origin of electron traps to oxygen vacancies, this might be linked with oxygen vacancies distribution across the dielectric energy band gap.

Detailed distributions for HfO_2 and HfAlO are shown in Fig.12a&b. Only two peaks are observed in HfO_2. The one at 0.45 eV above Si E_{CB} is dominant and the one near Si E_{CB} is much smaller. Only 10% of traps in HfO_2 are below Si E_{CB}. There are three peaks in the 1:1 cycle HfAlO, with peaks at Si E_{CB}, 0.45 eV above and -0.8 eV below Si E_{CB}, respectively. 70% of traps in HfAlO are above Si E_{CB}. The peak at the Si E_{CB} is much higher than that in HfO_2. The normalized energy distributions in HfO_2, HfAlO and Al_2O_3 are compared in Fig.13. The trap energy level in Al_2O_3 is clearly deeper than that in HfO_2, and HfAlO combines the features in HfO_2 and Al_2O_3. The deep peak at -0.8 eV in Al_2O_3 and HfAlO agrees well, which is missing in HfO_2 due to the absence of Al. HfO_2 has a large shallow peak at 0.45 eV, while Al_2O_3 has a much smaller one. The shallow peak in HfAlO is the average of that in HfO_2 and Al_2O_3. We can therefore conclude that Al induces the deep peak at -0.8 eV and Hf induces most of the shallow peak at 0.45 eV. This result demonstrates that DMP is capable of distinguishing the overall electron trap energy distribution across different high-k dielectric stacks, while the measured distributions for shallower traps agree well with CIS [8].

Recent retention test in memory cells confirms that higher Hf content results in a larger initial V_{th}-window closure, which can be attributed to the fast discharge from its larger amount of shallower traps [8]. The distributions obtained from several devices of the same stack, each with a different $T_{discharge}$ from 10^2 sec to 10^5 sec, essentially agree with each other, as shown in Fig.14, indicating that the trap energy distribution changes little with the location in the bulk of the high-k layer.

Conclusion

A new discharge-based multi-pulse technique has been developed. This work, for the first time, provides the overall electron trap energy distribution across the high-k dielectric layer. Trap energy distributions in HfO_2, Al_2O_3 and HfAlO have been extracted and compared. It is observed that hafnium is responsible for the shallow traps at about 0.45 eV above the Si E_{CB}, and aluminum causes the deep traps at -0.8 eV below the Si E_{CB}. HfAlO combines the features in HfO_2 and Al_2O_3. This technique can assist in explaining the retention behavior of floating gate Flash memory cells and help optimizing processes and materials.

References

[1] B. Govoreanu *et al, IEDM,* 2006.
[2] B. Govoreanu *et al, IEDM,* 2008.
[3] E. Cartier, *et al*, *IEDM,* 2006.
[4] M. B. Zahid *et al, IRPS,* 2007.
[5] Y. Wang *et al*, *IEDM,* 2006.
[6] R. Degraeve, *et al, IEDM,* 2008.
[7] W. Zhang, *et al*, *Elec. Dev. Lett., pp1043-1046,* 2008.
[8] B. Govoreanu, *et al, INFOS* 2009.

Table 1: Summary of the samples used in this paper. All devices are nMOSFETs. RTA at 900 ^{o}C was carried out for 30 s after the n+ poly deposition.

Stack	Process condition summary
1	1 nm SiO_2 /10 nm ALD HfO_2 with PDA@550 ^{o}C in N_2, poly gate.
2	2 nm SiON /10 nm ALD Al_2O_3 with PDA@700 ^{o}C in N_2, poly gate.
3	2 nm SiON /12 nm ALD HfAlO without PDA, poly gate.
4	2 nm SiON /10 nm ALD HfO_2 with PDA@500 ^{o}C in O_2, poly gate.

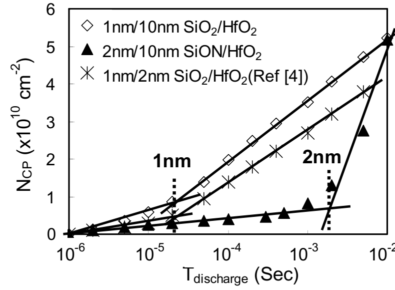

Fig.1 CP results on SiO_2/HfO_2 stacks with different SiO_2 thickness. The sharp increase in the slope at the dotted line is caused by transition from SiO_2 to HfO_2.

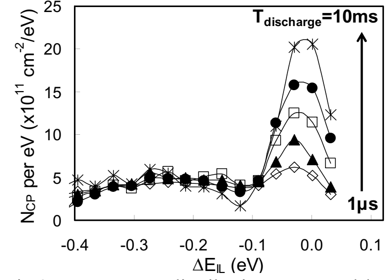

Fig.2 Trap energy distribution measured by CP on stack 1. Test conditions are V_{charge}=1.5V, V_{base}=-1.5~0.5V, t_{rf}=0.3 μs, T_{charge}=100 μs. ΔE_{IL} is the energy level with regard to Si conduction band bottom at flat band state.

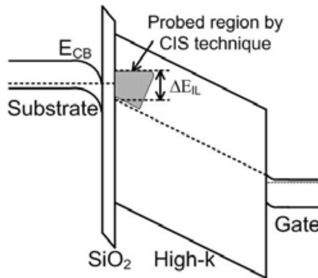

Fig.3 Energy band diagram illustrating that only a limited region above E_{CB} and within about 2~3 nm into the high-κ bulk can be probed by the CIS technique for a 1nm/12nm SiO_2/Al_2O_3 stack when the maximum charging time is 2000 seconds (ref.[6]).

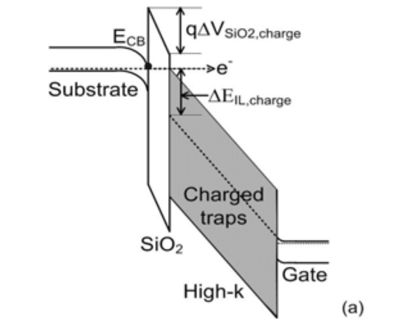

Fig.4a Illustration of energy band diagram of a SiO_2/high-k gate stack in an nMOSFET during the charging at a large positive gate bias. Electrons are injected into the conduction band of the high-κ layer to fill the traps through scattering.

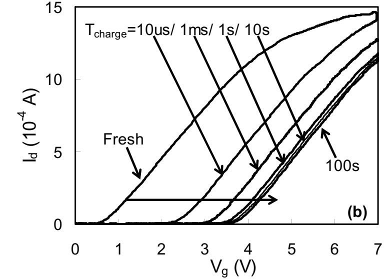

Fig.4b Id-Vg during charging the 2nm/10nm $SiON/Al_2O_3$ stack (stack 2) measured by pulsed technique. $V_{charging}$=7V.

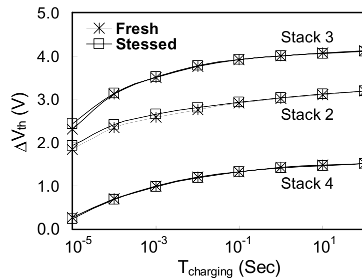

Fig.5 Trap charging kinetics on the fresh and the stressed-then-discharged device. Little trap generation is observed for the SiON/high-k stacks used in this work.

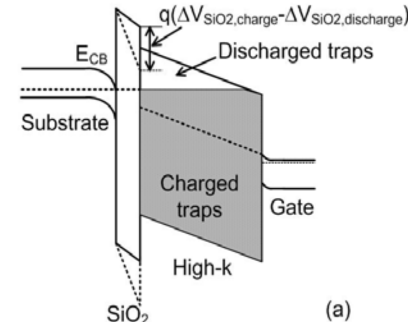

Fig.6a Illustration of energy band diagram of SiO_2/high-k gate stack in a nMOSFET during discharging at a low positive gate bias. Shallower traps in the white region of high-k layer can be discharged.

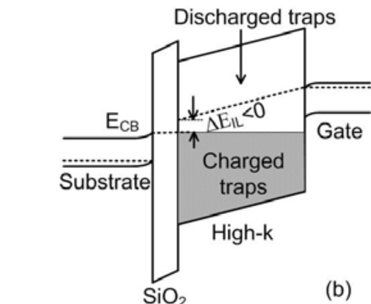

Fig.6b Illustration of energy band diagram of SiO_2/high-k gate stack in a nMOSFET during discharging at a negative gate bias. Deeper traps corresponding to ΔE_{IL}<0 are also discharged as their energy levels are increased to above the Si E_{CB} at the interface.

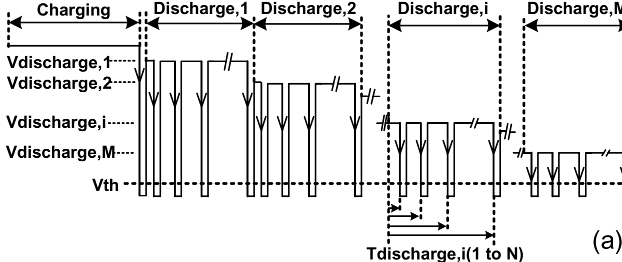

Fig.7a Illustration of the waveform for DMP technique when $V_{discharge}$>V_{th}. After charging, a number of lower voltage ($V_{discharge,1}$~$V_{discharge,m}$) is sequentially applied to discharge shallow traps. For each $V_{discharge}$, a new saturated trapping level is observed and the discharge is interrupted by short pulses of 5 μs edge time to measure V_{th}.

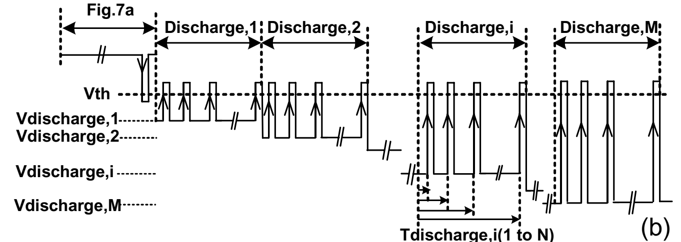

Fig.7b Illustration of waveform for DMP technique when $V_{discharge}$ is reduced below V_{th}. The short pulses change direction in order to measure V_{th} from the pulse edge. By combining Fig.7a&b, the energy distribution both below and above Si E_{CB} can be measured.

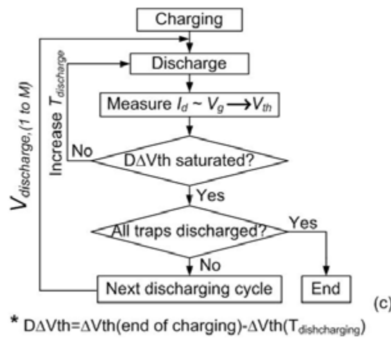

Fig.7c Flow chart of the test procedure. V_{th} measured at the end of each discharge stage is used as the initial V_{th} of the following discharge stage.

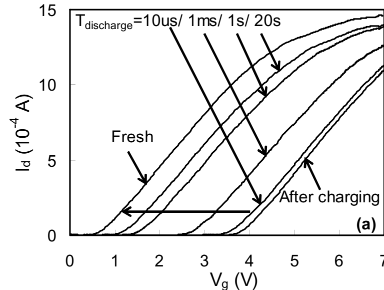

Fig.8a Id-Vg during the discharging of a n-MOSFET with 2nm/ 10nm SiON/Al_2O_3 gate stack (stack 2). V_{top}=7V, T_{charge}=100s, $V_{discharge}$=-5V, and the pulse rising/falling time on the edge is 5 µs.

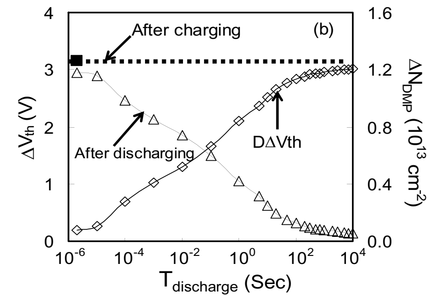

Fig.8b Trapping level after charging and discharging at different $T_{discharge}$, and their difference, DΔVth. The second Y-axis shows the equivalent discharged trap sheet density at the substrate/dielectric interface, ΔN_{DMP}. $V_{discharge}$=-5V.

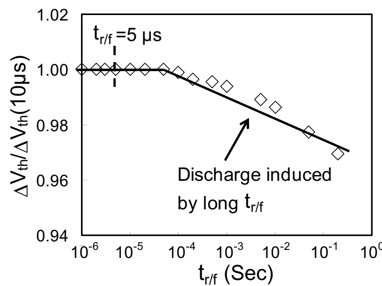

Fig.9 The effects of rising/falling time of the pulse on the measurement (stack 2). There is no additional discharge during measurement when using a 5 µs edge time.

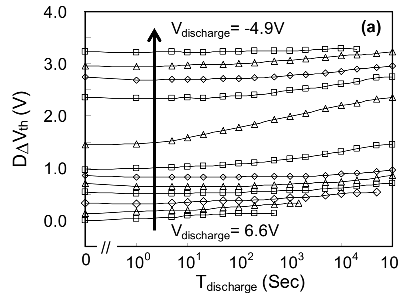

Fig.10a A typical result of DMP technique on an nMOSFET with stack 2. The device was charged at a gate bias of 7 V for 100 sec to fill the traps before the discharging.

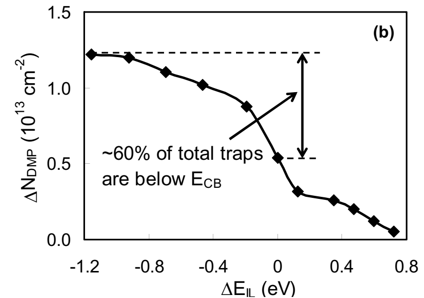

Fig.10b Cumulative trap density at the saturation point of each discharge vs. the corresponding energy level. Nearly 60% of the traps in Al_2O_3 are below Si E_{CB}.

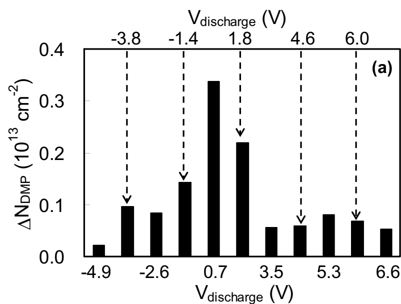

Fig.11a Trap density measured for each discharge stage vs. the corresponding $V_{discharge}$ for stack 2. The test conditions are the same as those in Fig.10.

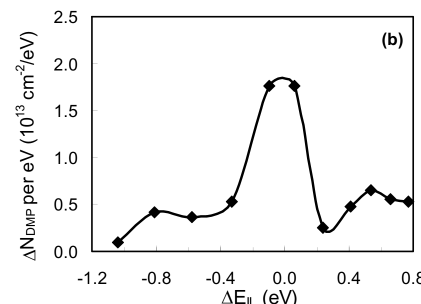

Fig.11b Trap density per eV vs. energy level for stack 2. Three peaks are observed at around Si E_{CB}, 0.5 eV above Si E_{CB}, and -0.8 eV below Si E_{CB} respectively. Only the peak at 0.5 eV can be measured by the CIS technique.

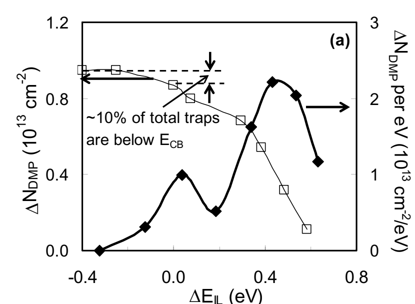

Fig.12a Cumulative trap density and trap density per eV vs. energy level in HfO_2 (stack 4). About 10% of the total traps in HfO_2 are below Si E_{CB}. Two peaks are observed at around Si E_{CB} and 0.45 eV above Si E_{CB}, respectively.

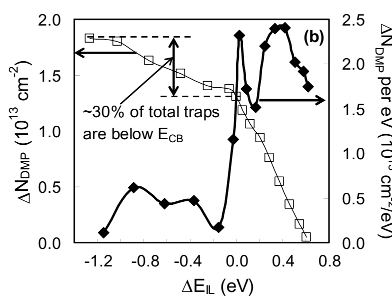

Fig.12b Cumulative trap density and trap density per eV vs. energy level in HfAlO (stack 3). About 30% of the traps in HfAlO are below Si E_{CB}. Three peaks are observed at around Si E_{CB}, 0.45 eV above E_{CB} and -0.9 eV below E_{CB} respectively.

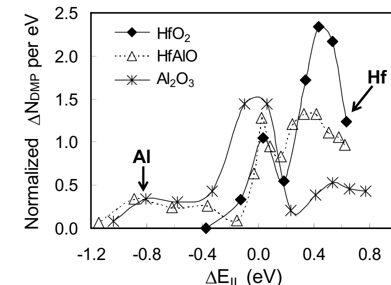

Fig.13 Comparison of the trap density per eV normalized by the total trap density in gate stacks with Al_2O_3 (stack2), HfAlO (stack3) and HfO_2 (stack4). The area underneath each curve is 1 after the normalization. Al induces the deep peak at -0.8 eV and Hf induces the shallow peak at 0.45 eV.

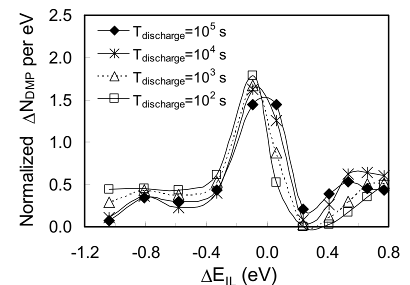

Fig.14 Normalized trap density per eV measured on several devices (stack 2) with different $T_{discharge}$ from 10^2 to 10^5 s, respectively. Trap energy distribution in the bulk of high-k layer changes little with the spatial location. Similar results were observed for all stacks.