<table>
<thead>
<tr>
<th><strong>Title</strong></th>
<th>Temperature-dependent relaxation current on single and dual layer Pt metal nanocrystal-based Al2O3/SiO2 gate stack</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Author(s)</strong></td>
<td>Chen, Y. N.; Goh, K. E. J.; Wu, X.; Lwin, Z. Z.; Singh, P. K.; Mahapatra, S.; Pey, Kin Leong</td>
</tr>
<tr>
<td><strong>Date</strong></td>
<td>2012</td>
</tr>
<tr>
<td><strong>URL</strong></td>
<td><a href="http://hdl.handle.net/10220/10897">http://hdl.handle.net/10220/10897</a></td>
</tr>
<tr>
<td><strong>Rights</strong></td>
<td>© 2012 American Institute of Physics. This paper was published in Journal of Applies Physics and is made available as an electronic reprint (preprint) with permission of American Institute of Physics. The paper can be found at the following official DOI: [<a href="http://dx.doi.org/10.1063/1.4764873">http://dx.doi.org/10.1063/1.4764873</a>]. One print or electronic copy may be made for personal use only. Systematic or multiple reproduction, distribution to multiple locations via electronic or other means, duplication of any material in this paper for a fee or for commercial purposes, or modification of the content of the paper is prohibited and is subject to penalties under law.</td>
</tr>
</tbody>
</table>
Temperature-dependent relaxation current on single and dual layer Pt metal nanocrystal-based Al2O3/SiO2 gate stack


Citation: J. Appl. Phys. 112, 104503 (2012); doi: 10.1063/1.4764873

View online: http://dx.doi.org/10.1063/1.4764873

View Table of Contents: http://jap.aip.org/resource/1/JAPIAU/v112/i10

Published by the AIP Publishing LLC.

Additional information on J. Appl. Phys.

Journal Homepage: http://jap.aip.org/
Journal Information: http://jap.aip.org/about/about_the_journal
Top downloads: http://jap.aip.org/features/most_downloaded
Information for Authors: http://jap.aip.org/authors
Temperature-dependent relaxation current on single and dual layer Pt metal nanocrystal-based Al$_2$O$_3$/SiO$_2$ gate stack

Y. N. Chen,$^{1, a)}$ K. E. J. Goh,$^2$ X. Wu,$^{1, 3}$ Z. Z. Lwin,$^1$ P. K. Singh,$^4$ S. Mahapatra,$^4$ and K. L. Pey$^{1, 3}$

$^1$Microelectronics Center, School of EEE, Nanyang Technological University, Nanyang Avenue, Singapore 639798, Republic of Singapore
$^2$Institute of Republic of Material Research and Engineering, A STAR (Agency for Science, Technology and Research), 3 Research Link, Singapore 117602, Republic of Singapore
$^3$Singapore University of Technology and Design, Singapore 279823, Republic of Singapore
$^4$Indian Institute of Technology Bombay, Powai, Mumbai 400076, India

(Received 18 January 2012; accepted 12 October 2012; published online 21 November 2012)

We present a systematic investigation of the temperature dependent relaxation current behavior for single layer and dual layer Pt metal nanocrystal (MNC)-based Al$_2$O$_3$/SiO$_2$ flash memory gate stacks. Stacks containing single layer Pt MNC exhibit a dual-slope behavior in the log-log plots of the relaxation transient, whereas those with dual layer Pt MNC exhibit a single-slope behavior. We propose a physical model embodying two competing relaxation mechanisms to explain the Pt MNC induced relaxation current—thermionic emission and the quantum tunneling. Based on this model, the dual-slope behavior of single layer MNC-based gate stack can be ascribed to the dominance of thermionic emission at the initial part and quantum tunneling at the tail part. In contrast, the single slope behavior of the dual layer metal nanocrystal-based stack arises from the dominance of the quantum tunneling throughout the relaxation. In addition, we verify that stacks containing dual layer MNC show better retention property than their single layer counterparts. Our results demonstrate that relaxation current measurements offer a simple way to assess the charge retention capability for MNC-based gate stacks.

I. INTRODUCTION

Continual scaling of traditional floating gate flash memory cell size below 22 nm is challenging because of the bottleneck issues for scaling tunnel oxide, lateral dimension, and control oxide.$^1$ A novel structure adopting a single layer (SL) or dual layer (DL) metal nanocrystal (MNC) as localized charge storage medium has been proposed as a candidate for sub-22 nm flash memory cell.$^2$ Besides the possibility of scaling below 22 nm, the discrete individual charge storage nodes (i.e., MNCs) also provide better immunity to defects than conventional floating gate memory.$^3$ In particular, it has been reported that stacks containing DL MNC provide an enhanced memory window over SL MNC and have the potential capability of storing multi-bits per cell for the future NAND flash applications.$^2$ Extensive reliability studies have been carried out for this type of novel memory cells.$^{2, 4-6}$ In this study, dielectric relaxation current measurement is used to characterize the electrical properties of flash memory gate stack containing SL and DL MNC with Al$_2$O$_3$ and SiO$_2$ layer as the respective control and tunnel oxide layers. A physical model is proposed to elucidate the origin of different behaviors of the relaxation current in the SL and DL Pt MNC gate stacks.

II. EXPERIMENTAL

The tunnel oxide for this MNC-based flash memory gate stack was prepared by thermally grown SiO$_2$ using in-situ steam-generation process in an applied materials 200 mm rapid thermal process (RTP) chamber. MNCs were formed by first depositing a thin film of Pt and then annealing it at high temperature for 30 s. The metal thin film would ball up into numerous MNCs (roughly uniform in size) because of surface energy minimization.$^7$ The control oxide for the SL MNC gate stack of Al$_2$O$_3$ (high-$k$ (HK)) is deposited by physical vapor deposition (PVD). For the DL MNC-based gate stack, after forming the bottom layer of MNC, a thin film of Al$_2$O$_3$ was deposited on top as the inter layer dielectric (ILD). This is followed by the formation of the top layer MNC. A thick layer of Al$_2$O$_3$ was finally deposited as the control oxide on top. A pure HK layer of 10 nm Al$_2$O$_3$ deposited directly to Si substrate (with approximately 1 nm interfacial layer (IL) SiO$_2$) was used in this study as a control sample. The top metal electrodes for all samples are Pt dots of 160 $\mu$m diameter and 100 nm thickness realized by shadow masking. Figures 1(a)–1(c) show the schematic of

![Fig. 1. Schematic cross-sections showing dimensions of three samples: (a) Control HK, (b) SL MNC-based gate stack, and (c) DL MNC-based gate stack.](image-url)

$^a$Email: E070048@ntu.edu.sg.
III. RESULTS AND DISCUSSION

A. Temperature-dependent \( J_{relax} \) slope of SL and DL MNC-based gate stack samples

When a sudden electric field change is applied to (or removed from, in our case) capacitors with high-\( \kappa \) dielectric gate stacks, the leakage current amplitude reaches a peak value immediately and then decays with time.\(^9\) This current is referred to as dielectric relaxation current, which is often observed in dielectric materials. A normal dielectric relaxation current follows the sign of \( dV/dt \) and follows the Curie-von Schweidler law: \( J_{relax} = at^n \),\(^10\) which is a straight line on a log-log scale with a slope \( n \). Consistent with previous reporting of the relaxation current by Reisinger \textit{et al.},\(^11\) \( n \) of a pure HK is a constant number close to one and is independent of temperature. It is also shown\(^11\) that the contribution of the total relaxation current from the SiO\(_2\) layer is negligible because of its small magnitude compared to that from the high-\( \kappa \) layer. Thus, in this study, we assume that the \( J_{relax} \) comes primarily from the high-\( \kappa \) layer (with or without MNC). It is also shown that the \( J_{relax} \) of the high-\( \kappa \) dielectric (such as HfO\(_2\), ZrO\(_2\), and Al\(_2\)O\(_3\)) is independent of temperature and proportional to the initial electric field \( (E_{ini}) \).\(^11\)

In this study, \( E_{ini} \) for the relaxation current measurement on the high-\( \kappa \) Al\(_2\)O\(_3\) layer (embedded with SL or DL MNCs) of the SL and DL sample was calculated by

\[
E_{ini} = \frac{(V_{ini} - V_{FB} - \Psi)}{t_{HK} \cdot EOT_{stack}^3},
\]

where \( V_{ini} \) is the magnitude of the initial step voltage applied to the top stack for the production of \( J_{relax} \), \( V_{FB} \) is the initial drain voltage of the MOS capacitor, \( \Psi \) is the initial Si surface potential, \( EOT_{HK} \) and \( EOT_{stack} \) are the equivalent SiO\(_2\) thickness of the total Al\(_2\)O\(_3\) layer and the whole gate stack, respectively, and \( t_{HK} \) is the thickness of the total Al\(_2\)O\(_3\) layer. In this study, all the \( E_{ini} \) are chosen to be the same value of 3 mV/cm for the ease of parallel comparison.

Figure 3 shows the \( J_{relax} \) of the control Al\(_2\)O\(_3\) sample (control HK) at 300, 350, 400, and 450 K, and the inset lists the slope \( n \) extracted from each curve fitting at the log-log scale. Consistent with previous reporting of the relaxation current,\(^11\) \( n \) of a pure HK is a constant number close to one and is independent of temperature. The amplitude of \( J_{relax} \) increases slightly with temperature (\( T \)). This is probably because of the increased charge trapping/de-trapping in the oxide layers with increasing \( T \).\(^12\) The \( J_{relax} \) of the SL Pt MNC sample at the same four temperatures is shown in Figure 4. It is evident that \( J_{relax} \) curve no longer shows a single slope but manifests primarily two parts with different slopes. These two parts of the \( J_{relax} \) exhibit different temperature dependencies. We consider as the “initial part” the data from the first data point up to 20 s and the “tail part” as the points from 40 s to the end of the measured time range. The slope \( n \) extracted from the curve-fitting of the two parts is shown in the two insets. It can be seen that \( n \) increases with increasing \( T \) in the initial part, whereas in the tail part \( n \) appears to decrease with \( T \).

In contrast, a different dependence of \( J_{relax} \) on temperature is observed for the DL MNC sample. As shown in Figure 5, \( J_{relax} \) shows significant fluctuation especially at elevated temperatures, which is also observed in the tail part of \( J_{relax} \) in
Figure 4. The inset lists the slope $n$ of the DL MNC sample from general curve-fitting. It is clear that $n$ decreases as $T$ rises, as if the DL sample exhibits only the behavior of the tail part of the SL sample.

B. Role of MNC in relaxation current

To compare the magnitude $J_{\text{relax}}$ of the control HK, SL MNC, and DL MNC samples, we shall focus on the trend of $J_{10}$ and $J_{60}$ as a function of $T$, as shown in Figure 6. It can be observed from Figure 6(a) that the magnitude of $J_{10}$ for the control sample is relatively insignificant compared to that of $J_{10}$ for the SL and DL Pt MNC-based gate stack samples, indicating that the dominant $J_{\text{relax}}$ behavior in the SL and DL samples came from the MNC, especially at the initial stage of relaxation. Hence, we neglect the effects in $J_{\text{relax}}$ induced by the Al$_2$O$_3$ layer in our analysis. The $J_{10}$ value of the SL sample is observed to decrease slightly with increasing $T$, in accordance with the increasing slope $n$. However, the $J_{10}$ value of the DL sample exhibits the opposite behavior, i.e., increases with rising $T$. Furthermore, the $J_{10}$ value of the DL shows much higher magnitude than the SL, especially at high temperatures.

Here, we propose a model to explain the $J_{\text{relax}}$ behavior of the SL and DL MNC-based high-$\kappa$SiO$_2$ gate stack at different temperatures. When the SL and DL MNC-based gate stack is under positively biased $V_{\text{ini}}$, as shown in the band diagram of Figure 7(a) and Figure 8(a), the electrons are being injected from the substrate. The electron flux direction is from substrate toward the top electrode, which is basically a programming process of the MNC-based flash memory gate stack. The electrons could be trapped by the potential wells resulting from the embedded MNCs (charging of the MNCs). The energy level needed to trap an additional electron in the MNC is described by

$$E_c = \frac{e^2}{2C},$$  \hspace{1cm} (2)

where $e$ is the electronic charge and $C$ is the self-capacitance of the charged MNC. The energy states of the charged electrons are quantized.

Upon the removal of the $V_{\text{ini}}$, a negative relaxation current is observed, indicating that the direction of the electron flux now comes from the top electrode towards the Si substrate. Since the magnitude of HK-induced relaxation current is relatively small compared to the MNC-induced relaxation current at the early stage of the $J_{\text{relax}}$ measurement (see Figure 6(a)), we shall focus primarily on the $J_{\text{relax}}$ contributed by the MNC. We propose that there are mainly two different trap levels in the MNC, shallow and deep traps. Thus, we consider that there are two main components of the MNC-induced $J_{\text{relax}}$. One part arises from the charges that are no...
longer confined in the MNC because of the sudden removal of $V_{ini}$. These charges are probably the trapped electrons that were at the top energy states of the potential wells of the MNCs (shallow trap electrons). As shown by the dashed line in Figure 7(b), it is proposed that these charges leak from the MNCs and travel to the Si substrate through thermionic emission process:  

$$J_1 = J_1(t) \propto T^2 \exp \left[ \frac{\epsilon}{kT} \left( a \sqrt{E_i(t)} d - \phi_B \right) \right],$$  

(3)  

where $J_1$ is the time-dependent thermionic emission current density, $\epsilon$ is the elementary electron charge, $T$ is the absolute temperature, $a$ is a constant, $E_i(t)$ is the time-dependent internal electric field across the SiO$_2$ tunnel oxide with a maximum value at $t = 0$, $d$ is the thickness of the tunnel oxide and $\phi_B$ is the barrier height of the MNC material. The internal electric field across the SiO$_2$ tunnel oxide is a result of trapped electrons in the potential well of MNC. At the moment $V_{ini}$ is removed ($t = 0$), there is certain number of trapped electrons, and as these electrons leak out of the MNC ($t > 0$), internal electric field would drop. So $E_i(t)$ and $J_1$ both decay with time.

The other component of the MNC-induced $J_{relax}$ is the charge leakage current that comes from the deep trap electrons tunneling through the SiO$_2$ tunnel oxide and reaching the Si substrate (solid line in Figure 7(b)). This charge leaking process is proposed to be quantum tunneling since there is no external electric field on the stack and the SiO$_2$ oxide layer is relatively thin  

$$J_2 = J_2(t) \propto [E_i(t)d]^2 \exp[-b/E_i(t)d],$$  

(4)  

where $J_2$ is the time-dependent quantum tunneling current density, $b$ is a constant, and $E_i(t)$ and $d$ have the same meanings as in Eq. (3). $E_i(t)$ and $J_2(t)$ also decay with time because of the fixed number of initial trapped electrons in the MNC. Both thermionic emission and quantum tunneling mechanisms could also result in electrons going toward the top electrode, but since only negative relaxation current is measured, these electrons may be annihilated or masked by the main electron flux.

Similar to the SL MNC sample, both thermionic emission current $J_1$ and quantum tunneling current $J_2$ comprise the $J_{relax}$ in DL MNC sample (see as shown in Fig. 8(b)). The difference is that, there is considerable probability that the electrons released by thermionic emission from one layer of MNC could be re-trapped by another layer of MNC in the DL-MNC based gate stack. On the other hand, the quantum tunneling could happen not only from the electrons of the bottom layer of MNC1 but also via the two-time tunneling of the electrons of the top layer of MNC2. The possibility of the two-time tunneling is discussed as follows. At equilibrium state, there is no net electric field in the ILD Al$_2$O$_3$, as shown in Figure 8(b). However, as the electrons of MNC1 tunnel out, the charge equilibrium of the two layers is broken and a subtle internal electric field across the very thin ILD with a direction from MNC1 to MNC2 will arise. This internal electric field results in a finite probability for the electrons from MNC2 to tunnel through the ILD to MNC1, until a new equilibrium is established. Although occurrence of the thermionic emission current between MNC1 and MNC2 is also possible, the random nature of this process is assumed to produce no net current contribution. Thus, the contribution of $J_1$ is less significant in the DL MNC-based sample and we expect that the contribution of $J_2$ to play a more dominant role in the total $J_{relax}$ for the DL sample than the SL sample.

When the MNC-based gate stack was under $V_{ini}$ before the relaxation current measurement, it is proposed that both thermionic emission current $J_1$ and quantum tunneling current $J_2$ contribute to the programming current. Equation (3) suggests $T$-dependence of the thermionic emission current $J_1$. As the carrier density of the Si substrate also increases with $T$, the corresponding increase in the substrate conductivity implies for a given $V_{ini}$ and stressing time, the programming current would be higher and more electrons get stored at the MNC (assuming they are not saturated). This implies that $E_i(t = 0)$ increases with $T$. This explains the slight increase in the first data point in the $J_{relax}$ curve as increasing $T$ in Figure 4. In Figure 5, the increase in $J_{relax}$ at the first data point is significant because of the increased number for electron-trapping centres associated with the MNC. We propose that, in the SL sample, at the early stage of the relaxation current measurement, as the external voltage is dropped to zero, and there are no more programming electrons flux, the band diagram of the SL MNC-based gate stack has changed from Fig. 7(a) to 7(b). There are considerable electrons de-trapping from the MNCs over the tunnel oxide via thermionic emission, i.e., $J_1$. Thus, at the initial part of the SL sample $J_{relax}$. $J_1$ resulting from the trapped electrons released from the MNCs dominates the behavior of $J_{relax}$. As $J_1$ increases dramatically with $T$, indicating that discharge rate of the electrons in the initial part of $J_{relax}$ increases remarkably. Thus, at a given time $t = t_1$, even though the total amount of the electrons to discharge increases with $T$, there are less “remaining electrons” in the SL MNC-based gate stack with increasing $T$ (the “remaining electrons” for the gate stack to discharge during the relaxation process could roughly be the area covered by the $J_{relax}$ ranging from $t_1$ to $t_{1,relax} = \infty$). So, as $J_1$ dominates the initial part of $J_{relax}$, a decrease in the $J_{relax}$ value in Figure 6(a) and a increasing slope $n$ of the initial part of $J_{relax}$ in Figure 4 with increasing $T$ were observed.
As the $J_1$ electrons leak out, the magnitude of $J_{\text{relax}}$ drops dramatically until a transition happens where $J_2$ becomes dominant in the tail part of the $J_{\text{relax}}$. Equation (4) suggests no $T$ dependence of $J_2$, so $J_2$ should decay with time at a fixed rate at all four temperatures. However, because of the increased programming current which results in extra amount of trapped electrons in the MNC, the total number of the trapped electrons increases with $T$. Thus, at certain time of the tail part of the relaxation, the remaining charge in the gate stack increases with $T$, leading to a decreasing slope $n$ in Figure 4 and a slightly increasing $J_{\text{60s}}$ in Figure 6(b).

In the case of DL MNC-based gate stack sample, as the $J_1$ component of $J_{\text{relax}}$ is not as significant in the DL sample, we propose that the behavior of $J_{\text{relax}}$ mainly arises from $J_2$ (no $T$ dependence). In addition, there is an increased amount of trapped electrons with increasing $T$ because of an extra layer of MNCs. Thus, we can deduce that the residual charge in the DL MNC sample increases with $T$. As a result, a more significant decreasing of slope number $n$ in Figure 5, as well as the increase in $J_{\text{10s}}$ and $J_{\text{60s}}$ in Figures 6(a) and 6(b) with $T$, are observed in DL MNC sample.

It was observed that at elevated temperatures, there is increased fluctuation of the $J_{\text{relax}}$ value in the DL sample. These fluctuations could arise from two origins: thermal noise and trapping-detrapping of the electrons. The thermal noise could be from the background, the probe station, the semiconductor characterization system, as well as the MNC-based gate stack itself. Furthermore, it is proposed that an increased number of charge trapping-detrapping events occur through thermionic emission between MNC1 and MNC2 happen at elevated temperature, contributing to the fluctuations.

C. Relaxation current as a reliability tool

As indicated in Figures 6(a) and 6(b), the main part of the relaxation current of the MNC-based gate stack comes from the MNC, because of the insignificance of the $J_{\text{10s}}$ and $J_{\text{60s}}$ values of the control HK. As the MNC-induced relaxation current is basically the short-term leakage current of the trapped-charge in the MNC, the relaxation current can be used as a reliability tool to evaluate the short-term retention properties of the MNC. Here, we define a discharge time constant $\tau$ as the time taken for the relaxation current to decay to 37% (or 1/e) of its initial value.$^{16}$ The value of $\tau$ could be extracted from Figures 4 and 5. This parameter provides an indication of the short-term charge retention property of MNC-based flash memory cells. The extracted values of $\tau$ are listed in Figure 9 for the SL and DL MNC-based gate stack samples at four temperatures. Because of the different dominance in the relaxation current in the SL and DL samples discussed above, the $\tau$ values show different temperature dependencies. It is noticed that at the temperature of 300 K, the value of $\tau$ of the DL sample is slightly larger than the SL and control HK samples. A much larger $\tau$ value is observed for the DL sample at high temperatures than the SL, indicating a better short-term charge retention performance of the DL MNC-based high-$\kappa$/SiO$_2$ gate stack. This supports the enhancement of charge retention behavior by introducing an extra layer of MNC.$^2$

IV. CONCLUSION

We have investigated the relaxation current behavior of high-$\kappa$/SiO$_2$ flash memory gate stacks embedded with SL and DL MNC at different temperatures. We observed a unique dual slope behavior in the log-log plot of the relaxation current transient as a function of discharge time for the SL MNC sample but correspondingly only a single slope behavior for the DL MNC-based gate stack. We propose a model to explain these behaviors based on a competition between the thermionic emission current and quantum tunneling current. Using this model, the dual-slope behavior for the SL sample arises from the respective dominance of the thermionic emission in the initial part and quantum tunneling in the tail part of the relaxation current. Whereas the single slope behavior of the DL sample is mainly attributed to the dominance of the quantum tunneling throughout the relaxation process. In addition, we also extracted a discharge time constant from the relaxation current which provides support for DL MNC-based gate stacks exhibiting better short-term charge retention property, especially at high temperatures.

ACKNOWLEDGMENTS

This project is supported by the Singapore Ministry of Education research Grant No. T206B1205 and A*STAR thematic research Grant No. 042 114 0051.

NAND application,” in IEEE International Memory Workshop, IMW ’09 (IEEE, 2009), p. 4.


