NITRIDE ENGINEERING AND THE EFFECT OF INTERFACES ON CHARGE TRAP
FLASH PERFORMANCE AND RELIABILITY


aIndian Institute of Technology, Bombay, Mumbai, India
bApplied Materials, Santa Clara, USA; E-mail: sandhyac@ee.iitb.ac.in

ABSTRACT

The performance and reliability of Charge Trap Flash with single and bi-layer Si-rich and N-rich nitride as the storage node is studied. Single layer devices show lower memory window and poor cycling endurance, and the underlying physical mechanisms for these issues are explained. An engineered trap layer consisting of Si-rich and N-rich nitride interfaced by a SiON barrier layer is proposed. The effect of varying the SiON interfacial layer position on memory window and reliability is investigated. Optimum bi-layer device shows higher memory window and negligible degradation due to cycling (at higher memory window) compared to single layer films. The role of SiON interface in improving the performance and reliability of bi-layer stacks is explained.

[Keywords: SONOS EEPROMs, Charge Trap Flash, Bandgap Engineering, Endurance, Data Retention]

INTRODUCTION

Over the last few years, NAND Flash memories are experiencing tremendous growth because of data storage applications for portable electronics. As evident from the ITRS roadmap, NAND Flash scaling is progressing at a much faster rate than CMOS logic and is likely to hit the “red brick wall” in very near future [1]. It has been shown that Conventional Floating Gate (FG) Flash is difficult to scale beyond the 3X node due to cell to cell interference, loss of Control Gate to FG coupling, reduction in FG volume (hence the number of electrons stored per bit) [2] and the inability to scale the tunnel oxide thickness below about 7-8nm [1]. Of all the possible alternatives, Charge Trap Flash (CTF) [3] is an attractive candidate as it exhibits negligible cell to cell interference, planar structure hence better scalability and fully CMOS compatible fabrication process. However, CTF suffers from conflicting trends in memory window versus data retention and needs careful attention before becoming a viable technology option.

In the past, band gap engineering of the tunnel dielectric [4], [5] as well as of the silicon nitride storage layer [6], [7] has been done to improve CTF memory window and reliability. The use of multi-layer deposited tunnel dielectric [4], [5] instead of conventional thermally grown SiO2 may compromise the overall cell reliability, especially for Multi-Level Cell (MLC) operation that requires high Write/Erase (W/E) voltages for high memory window. Therefore, engineering the silicon nitride charge trap layer is a promising strategy to enhance the performance of CTF. Nitride composition gradients that resulted in a tapered band-gap structure demonstrated high memory window and retention but poor cycling endurance [6]. Recently, the use of silicon nitride-aluminum oxide-silicon nitride (NAN) composite as a storage node shows improved performance and reliability over single uniform nitride layer (this work) and graded-nitride devices reported elsewhere [6]. Comparable performance to the advanced NAN stack [8] is demonstrated, though with the advantage of simpler process integration by avoiding the incorporation of high-k dielectrics. The physical mechanism behind performance and reliability enhancement and stack design tradeoffs are discussed.

EXPERIMENTAL DETAILS

The fabrication process flow of the SONOS structures used in this work is illustrated in Table-1. The gate stacks consist of a 3.5nm Rapid Thermal Oxide (RTO) as a tunnel dielectric, an LPCVD silicon nitride of varying thickness and composition as the charge trap layer, and a 8nm High Temperature Oxide (HTO) as a blocking dielectric.

TABLE 1. FABRICATION PROCESS SEQUENCE FOR SONOS FLASH CAPACITOR STRUCTURES

- N substrate wafer clean
- Tunnel oxide (3.5nm), RTO
- LPCVD nitride storage layer (4/6/8nm)
- HTO deposited top oxide (8nm)
- Post HTO anneal
- Poly deposition (10nm)
- Boron implant
- Metal contact formation
- Backside metallization (Al)

The Si2H6/NH3 flow rates were varied during the LPCVD process to obtain Si-rich (Si+) and N-rich (N+) single layers having varying thicknesses (devices D1 through D5 as listed in Table-2). The Refractive Indices of N+ and Si+ blanket films were measured using a spectroscopic ellipsometer and was found to be 1.983 and 2.125 respectively, consistent with their compositional difference [9]. In bi-layer stacks, Si+ nitride of different thicknesses (see Table-2, devices D6 through D8) was deposited first, and then partially oxidized to produce ~2nm of SiON interface before N+ nitride deposition [10]. The position of the SiON interfacial layer depends on the initial Si+ nitride thickness as evident in Table-2. N2 anneal was done after HTO deposition, which was followed by poly deposition, implant and activation, gate metallization and gate patterning to complete device fabrication.

The device schematic, trap density and band diagram of single layer nitride stacks are illustrated in Fig. 1. The trap density of Si+ and N+ layers [11] and the band gaps [6] are indicated in accordance with the existing literature. Similar diagrams for the bi-layer nitride stacks with the interfacial SiON layer is indicated in Fig. 2. The narrow SiON interfacial layer, as depicted with lower trap density [12] and higher barrier height [9] in Fig. 2, is consistent with literature. The different bi-layer stacks (D6 through D8 in Table-2) have a progressively varying position of the SiON interface, which is essential to explore optimum interfacial position for best memory performance and reliability. The measurements were conducted on SONOS capacitor squares of length 100µm.
while the corresponding saturation $V_{FB}$ of N$^+$ nitride device is 7.5V at the same programming bias. Lower saturation $V_{FB}$ for Si$^+$ nitride can be linked to relatively shallow trap depth [14] in these films, which is consistent with published results [7], [11]. Note that the initial $V_{FB}$ changes while the saturation $V_{FB}$ remains unchanged after

only one W/E cycling. Therefore, the effective memory window for Si$^+$ nitride increases (3.3V to 4V) while that for N$^+$ nitride devices reduces (5.5V to 4V) after just a single W/E cycling at identical W/E bias (19V/16V).

Figs. 5 and 6 respectively show the erase transients for Si$^+$ and N$^+$ nitride SONOS devices after programming to different $V_{FB}$ levels by using different programming biases. Irrespective of programmed $V_{FB}$ level, the devices can be erased to a final value, which however is much higher for N$^+$ nitride than that obtained for Si$^+$ nitride. While over erase is observed for Si$^+$ nitride, the N$^+$ nitride devices show evidence of permanent trapping (results in $V_{FB}$ shift of ~1.5V over the fresh device) as they cannot be erased (no matter how long the erase pulse is applied) below 3.5V as shown in Fig. 6. To explore this permanent trapping component, W/E cycling was performed at different W/E bias and on SONOS devices having different N$^+$ layer thickness. The extracted permanent trapping component as a function of N$^+$ layer thickness is shown in Fig. 7. Note that a non-erasable component of ~1.5V is observed in all cases irrespective of the N$^+$ layer thickness. This suggests the occurrence of permanent trapping at the blocking-HTO/N$^+$ nitride interface. It is important to note that Si$^+$ nitride devices never show this permanent trapping component irrespective of the Si$^+$ nitride thickness. Therefore in spite of high saturation $V_{FB}$, poor erase performance reduces the useable memory window for N$^+$ nitride SONOS devices.

### RESULTS AND DISCUSSION

#### Single Layer Stacks

Figs. 3 and 4 respectively show the program transients for Si$^+$ and N$^+$ nitride SONOS devices at different biases. The transients were measured on fresh devices (open symbols) as well as after one W/E cycle (closed symbols). Some important differences between the two devices can be readily observed. First, fresh Si$^+$ nitride devices show lower fixed charge and a starting $V_{FB}$ of 0.2V than fresh N$^+$ nitride (starting $V_{FB}$ of 2V). After one W/E cycle, the fixed charge for Si$^+$ nitride reduces as starting $V_{FB}$ reduces to ~0.5V and over erase is observed, while that for N$^+$ nitride increases as starting $V_{FB}$ increases further to ~3.5V. The Si$^+$ nitride devices show faster programming with the onset of $V_{FB}$ saturation at ~1ms, as opposed to slower programming and onset of saturation at ~1s as seen for N$^+$ nitride. Faster program observed in Si$^+$ nitride is consistent with higher trap density reported in such films [13]. However, Si$^+$ nitride shows much lower saturation $V_{FB}$ of 3.5V at 19V programming bias, while the corresponding saturation $V_{FB}$ of N$^+$ nitride device is 7.5V at the same programming bias. Lower saturation $V_{FB}$ for Si$^+$ nitride can be linked to relatively shallow trap depth [14] in these films, which is consistent with published results [7], [11]. Note that the initial $V_{FB}$ changes while the saturation $V_{FB}$ remains unchanged after

only one W/E cycling. Therefore, the effective memory window for Si$^+$ nitride increases (3.3V to 4V) while that for N$^+$ nitride devices reduces (5.5V to 4V) after just a single W/E cycling at identical W/E bias (19V/16V).

Figs. 5 and 6 respectively show the erase transients for Si$^+$ and N$^+$ nitride SONOS devices after programming to different $V_{FB}$ levels by using different programming biases. Irrespective of programmed $V_{FB}$ level, the devices can be erased to a final value, which however is much higher for N$^+$ nitride than that obtained for Si$^+$ nitride. While over erase is observed for Si$^+$ nitride, the N$^+$ nitride devices show evidence of permanent trapping (results in $V_{FB}$ shift of ~1.5V over the fresh device) as they cannot be erased (no matter how long the erase pulse is applied) below 3.5V as shown in Fig. 6. To explore this permanent trapping component, W/E cycling was performed at different W/E bias and on SONOS devices having different N$^+$ layer thickness. The extracted permanent trapping component as a function of N$^+$ layer thickness is shown in Fig. 7. Note that a non-erasable component of ~1.5V is observed in all cases irrespective of the N$^+$ layer thickness. This suggests the occurrence of permanent trapping at the blocking-HTO/N$^+$ nitride interface. It is important to note that Si$^+$ nitride devices never show this permanent trapping component irrespective of the Si$^+$ nitride thickness. Therefore in spite of high saturation $V_{FB}$, poor erase performance reduces the useable memory window for N$^+$ nitride SONOS devices.

### TABLE 2. SPLIT CONDITIONS OF THE SONOS DEVICES WITH DIFFERENT NITRIDE CHARGE STORAGE LAYER COMPOSITIONS, 3.5nm TUNNEL OXIDE AND 8nm BLOCKING OXIDE.

<table>
<thead>
<tr>
<th>Split</th>
<th>Nitride layer type</th>
<th>Thickness (nm)</th>
<th>Measured EOT (nm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>D1</td>
<td>Si</td>
<td>4</td>
<td>13.06</td>
</tr>
<tr>
<td>D2</td>
<td>Si</td>
<td>6</td>
<td>14.20</td>
</tr>
<tr>
<td>D3</td>
<td>N</td>
<td>4</td>
<td>13.24</td>
</tr>
<tr>
<td>D4</td>
<td>N</td>
<td>6</td>
<td>14.67</td>
</tr>
<tr>
<td>D5</td>
<td>N</td>
<td>8</td>
<td>15.86</td>
</tr>
<tr>
<td>D6</td>
<td>Si bottom/ N top</td>
<td>2/6</td>
<td>15.22</td>
</tr>
<tr>
<td>D7</td>
<td>Si bottom/ N top</td>
<td>4/4</td>
<td>15.89</td>
</tr>
<tr>
<td>D8</td>
<td>Si bottom/ N top</td>
<td>6/2</td>
<td>14.81</td>
</tr>
</tbody>
</table>

### FIG. 3. PROGRAM TRANSIENTS OF 6nm THICK Si$^+$ NITRIDE SHOWING HIGHER MEMORY WINDOW AFTER THE FIRST W/E OPERATION.

### FIG. 4. PROGRAM TRANSIENTS OF 6nm THICK N$^+$ NITRIDE SHOWING A SMALLER MEMORY WINDOW AFTER THE FIRST W/E OPERATION.
Fig. 8 shows the programmed state retention characteristics of Si+ and N+ nitride SONOS devices at room temperature. The Si+ nitride device shows much larger VFB decay (0.6V in 104s) compared to the N+ nitride device, the later exhibiting a corresponding decay of only 0.04V. This suggests relatively (compared to N+ nitride) shallow trap depth in Si+ nitride, which is consistent with published results [7],[11], and justifies low saturation VFB and faster erase observed in such films.

Higher trap depth in N+ nitride is responsible for the excellent retention characteristics but makes erase slow in such devices.

Fig. 9 shows the W/E cycling endurance characteristics of N+ and Si+ nitride SONOS devices having varying charge-trap layer thicknesses. While the memory window is maintained after 10⁶ W/E cycles, a consistent increase in the erase-state VFB is observed for all devices indicating trap generation. Interestingly, the erase state VFB degradation (~0.8V) does not scale with nitride thickness irrespective of Si+ and N+ nitride stacks, indicating (from a simple electrostatics viewpoint [15]) the possible degradation (due to trap generation) of the blocking HTO/nitride interface.

Therefore, all single layer Si+ and N+ nitride SONOS devices show poor cycling endurance. The Si+ nitride devices show faster programming, no fixed charge problem and good erase, but suffer from low saturation VFB and poor retention. The N+ nitride devices on the other hand, show excellent retention and good saturation VFB, but show slow W/E speed and suffer from fixed charge issues.

Bi-layer Stacks

Fig. 10 shows the program transients of SONOS devices having bi-layer nitride stacks and compares them against that of a single layer N+ nitride device having equivalent thickness. Note that the initial fixed charge, believed to be due to permanent trapping at the blocking HTO/nitride interface (starting VFB) decreases markedly with the reduction in thickness of the top N+ layer. Furthermore, the saturation VFB also reduces with increasing bottom Si+ thickness. Note that the Si+/N+ = 2/6 (nm) bi-layer stack shows high permanent trapping while the Si+/N+ = 6/2 (nm) stack shows low saturation VFB, consistent with single layer results. The Si+/N+ = 4/4 (nm) bi-layer stack shows low permanent trapping and high saturation VFB and hence highest memory window. Fig. 11 shows the erase transients of
all bi-layer stacks and compares them with the N⁺ nitride single layer device. Note that while the single layer and Si⁺/N⁻ = 2/6 (nm) bi-layer stack are difficult to erase (show permanent trapping), the Si⁺/N⁻ = 4/4 and 6/2 (nm) devices show excellent erase capability. A comparison of saturated memory windows resulting from different programming voltages for various splits is shown in Fig. 12. All bi-layer stacks show a definite memory window improvement over single layer devices. The best memory window with low fixed charge, high saturation $V_{FB}$, high program speed and complete erase is demonstrated by the Si⁺/N⁻ = 4/4 (nm) bi-layer device.

Fig. 13 shows the retention loss from programmed state for all bi-layer stacks, and compares them to that for single N⁺ nitride device. The retention loss is sensitive to the position of SiON interface, and worsens with increasing bottom Si⁺ layer thickness for the bi-layer devices. The Si⁺/N⁻ = 2/6 (nm) stack shows negligible loss like the N⁺ nitride single layer stack. However the charge loss increases (shows a sudden jump) for the Si⁺/N⁻ = 4/4 (nm) device, and the Si⁺/N⁻ = 6/2 (nm) stack shows worse loss. To determine whether dominant charge loss is via the blocking top dielectric to gate or via the tunnel oxide to channel, retention was measured under varying gate bias stress as shown in Fig. 14. A positive gate bias stress would move the charge centroid towards the top dielectric, eventually resulting in charge loss by tunneling via the blocking dielectric to the gate. A negative gate bias stress on the other hand would push the electrons towards the tunnel dielectric and eventually enhancing the tunneling component via the bottom oxide. Note that charge loss was found to increase with negative bias, indicating that retention loss is primarily through the tunnel oxide. Therefore, the use of a slightly thicker tunnel oxide would reduce the retention loss observed in these devices.

Fig. 15 shows the W/E cycling endurance characteristics of all the bi-layer stacks. Compared to the single layer devices (see Fig. 9), all bi-layer devices show remarkable improvement in cycling induced erase state $V_{FB}$ degradation. Though there is a strong dependence on the position of the SiON interface layer, the erase state $V_{FB}$ shows negligible degradation for Si⁺/N⁻ = 4/4 and 6/2 (nm) bi-layer device. This is in stark contrast to the single layer nitride endurance results in this study as well as in the literature [3], [6], [7], [16], [17] and consistent with recent results on NAN stacks [8]. The importance of the SiON interfacial layer position in improving the...
Cycling endurance is worth a mention as the improvement is inexplicable by attribution to simply the ratio of Si+ vs. N+ nitride layer thickness in the bi-layer stack (as plausible in the case of W/E or retention performance). The signature effect of the SiON barrier layer explains the improvement for the bi-layer endurance as single layer nitride endurance is poor - regardless of the choice of nitride composition. The physical mechanism responsible for the differences between single layer and bi-layer stacks is discussed next.

**Physical Mechanisms**

As explained previously, the program speed and retention are qualitatively consistent with trap density and trap depth for single layer Si+ and N+ nitride films. Si+ nitride reportedly has higher trap density and lower trap depth [11] and therefore shows high program speed, good erase capability but low saturation VFB. The resultant higher charge mobility (hopping conduction) can explain the low fixed charge and no permanent trapping at the nitride/HTO interface in these films. The converse (low trap density and high trap depth) is true for N+ films, which manifests as large permanent trapping, good retention but high saturation VFB. Both Si+ and N+ nitride stacks show erase VFB degradation during cycling as a result of trap generation at the blocking HTO/nitride interface.

For bi-layer stacks, the potential barrier provided by the SiON interface (Fig. 1c) divides the trap layer into two potential wells and likely confines the injected charges to the bottom Si+ nitride potential well. This key insight can be obtained by comparing the erase VFB for single and bi-layer stacks (Fig. 9 and 15). As shown before, the HTO blocking oxide/N+ nitride interface is responsible for permanent charge trapping in the single layer devices. While this interface is also present in the bi-layer devices, it does not affect the erase state because carriers are blocked by the SiON barrier layer from reaching the N+ nitride/HTO interface and permanent trapping is avoided. This is evident from the dramatic fixed charge reduction as seen in Fig. 10 for the Si+/N+ = 4/4 and 6/2 (nm) devices, which also show excellent (like Si+ single layer) erase characteristics. It is interesting to note that the Si+/N+ = 2/6 (nm) bi-layer device still shows some permanent trapping and poorer erase as the bottom Si+ nitride layer gets consumed into SiON and the barrier is nonexistent.

The large discontinuity in retention as seen in Fig. 13 for bi-layer stacks having non-negligible bottom Si+ nitride is consistent with the material dependent trap depth in the vicinity of the critical charge loss path (i.e. the tunnel oxide) [6], [7], [12], [18], [19]. The composition independent nitride/HTO interface degradation with cycling in single layer nitride splits is avoided in Si+/N+ = 4/4 and 6/2 (nm) bi-layer nitride stacks while the Si+/N+ = 2/6 (nm) device shows degradation equal to single layer nitrides. This provides an interesting insight to trap generation mechanism and control in CTF devices with SiON interface layer position. The trap generation efficiency (number of generated traps per number of carriers traversing the gate stack during W/E [19]) gets reduced as charges tunneling at high electric fields are strongly scattered by the SiON interface barrier layer and thermalized to low energy carriers that reach the nitride/HTO interface and are unable to generate traps. The emergence of poor cycling endurance for the Si+/N+ = 2/6 (nm) stack as exhibited in Fig. 15, is possibly because the Si+ layer is converted to SiON, the bottom Si+ nitride potential well disappears and the scattering interfaces of the potential well are removed. This ineffectiveness of the SiON barrier (relative to Si+/N+ = 4/4 and 6/2 (nm) bi-layer nitride stacks) is consistent with the permanent trapping observed in this stack where the carriers populate the top (N+) nitride potential well in the absence of the bottom (Si+) potential well due to the position of the SiON interface layer.

**CONCLUSION**

Charge Trap Flash with single layer Si+ or N+ nitride as the storage node displays contrasting properties and shows clear performance and reliability trade-offs. The Si+ nitride exhibits high density of relatively shallow traps. In contrast, N+ nitride has deeper but fewer traps. Due to low saturation VFB for Si+ and high starting VFB for N+, overall memory window remains small for both stacks. N+ nitride show better retention but poorer erase capability compared to Si+ nitride. Both stacks show erase VFB degradation during cycling.

In comparison, bi-layer Si+/N+ nitride sandwiched stacks with an interfacial SiON layer show higher memory window and better cycling endurance. The SiON barrier can be optimally positioned to confine the carriers in the bottom Si+ nitride well. This results in simultaneous improvement in memory window and endurance. Trap generation in blocking HTO/nitride interface that causes erase VFB degradation during cycling is dramatically reduced by interface barrier insertion. The barrier helps reduce energy of electrons impinging on the blocking HTO/nitride interface and therefore reduces trap generation during cycling. Retention loss is attributed to tunneling of carriers through the tunnel dielectric. A thicker tunnel oxide can independently improve retention (independent of memory window and cycling endurance) to produce an overall optimized high performance CTF device.

**ACKNOWLEDGEMENT**

IIT Bombay would like to acknowledge SRC (GRC) for financial support.

**REFERENCES**


