Jurnal Teknologi, 38(D) Jun. 2003: 125–136

ONO AND TUNNEL OXIDE CHARACTERIZATION AND OPTIMIZATION FOR HIGH SPEED EEPROM DEVICE

UDA HASHIM[1] & RAMZAN MAT AYUB[2]

Abstract. Non-volatile memory processes, in particular the EEPROM process, is one of the hardest processes to be developed and the performance of the NVM products is normally judged from the programming speed and the density of the memory. The programming speed of the EEPROM cell depends critically on Tunnel Oxide Thickness (X_{tun}), Programming Voltage (V_p), ONO Thickness (X_{pp}) and Poly to Poly overlap Area (A_{pp}). However, in this experiment only ONO and tunnel oxide layer are optimized and characterized. Three experiments were setup to improve the programming speed. The first experiment was to scale down the ONO layer thickness and followed by measurement of the threshold voltage and breakdown voltage of the new ONO thickness. The second and third experiments were setup to check the integrity of ONO and tunnel oxide layers respectively. The EEPROM cell was fabricated to observe the cross sectional of ONO and tunnel oxide layer. The characterization work on ONO and tunnel oxide layer to increase the programming speed of the memory cells of a 16k EEPROM device has been carried out. After scaling down the nitride of ONO layer from 160? to 130?, the Vt program windows are further improved from 4.3V to 4.5V and from –0.7V to –0.9V for program high and program low operations, respectively. In this experiment, 130? was found to be the best thickness for nitride of ONO layer. The breakdown voltage for ONO at 130? of nitride thickness is 16.3V. The experiment revealed that the yields of ONO and tunnel oxide layer of the actual size on silicon were achieved at 98.7% and 99.92%, respectively.

Keywords: ONO, tunnel oxide, EEPROM, threshold voltage, programming speed, polysilicon, control gate, floating gate and select gate

1.0 INTRODUCTION

EEPROM is classified as a non-volatile memory (NVM) product; it refers to Electrically Erasable and Programmable Read Only Memories, and is available in two types, ie embedded and stand-alone [1]. Compared to a CMOS process, which only has 2 devices i.e. NMOS and PMOS, an EEPROM chip, is constructed from 6 devices, namely the low voltage NMOS (LVNMOS), the low voltage PMOS (LVPMOS), the high voltage NMOS (HVNMOS), the high voltage PMOS (HVPMOS), the natural NMOS (ZMOS), and the memory cell (EPROM cell).

[1] School of Microelectronics Engineering, Northern Malaysia University College of Engineering, Taman JKKK, Kubang Gajah, 02600 Arau, Perlis.
[2] NVM Research Group, MIMOS Semiconductor, MIMOS Berhad, Technology Park Malaysia, 57000 Kuala Lumpur, MALAYSIA, Email: uda@mimos.my

The performance of the NVM products is normally judged from the programming speed and the density of the memory [2, 3]. Besides conventional device characteristics, nonvolatile memory cells also have additional functional memory characteristics, which are used to evaluate the performance of the memory cell. One of the most important parameters is programming speed versus programming voltage, which describes the time dependence of the threshold voltage with respect to programming time, at different programming voltages.

A typical memory cell layout, called the Floating Gate Tunnel Oxide cell (FLOTOX) is shown in Figure 1 and the cross-section of the memory cell device is illustrated in Figure 2. In this experiment, 16k EEPROM is fabricated with a double-poly, double-

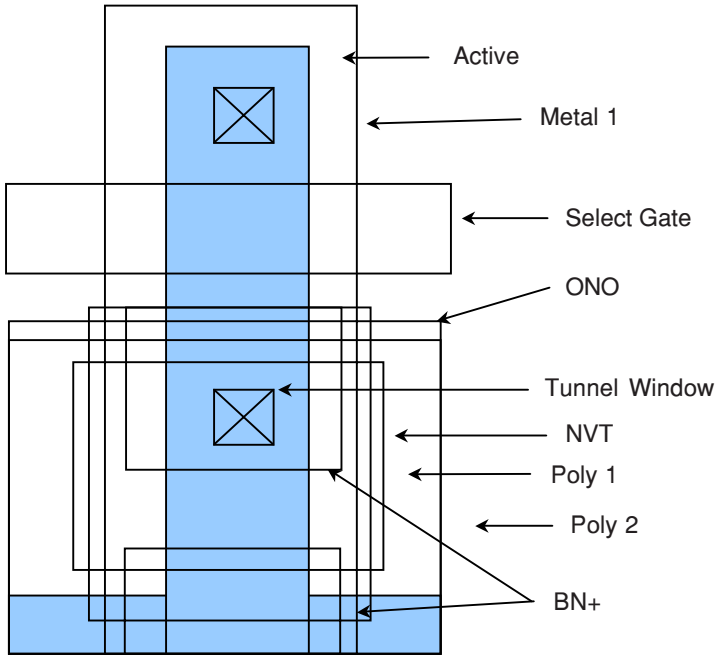


Figure 1 Typical of Memory cell layout

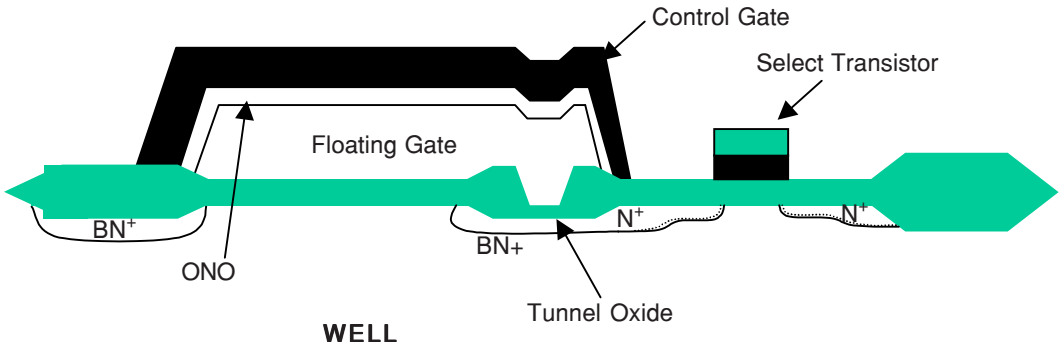


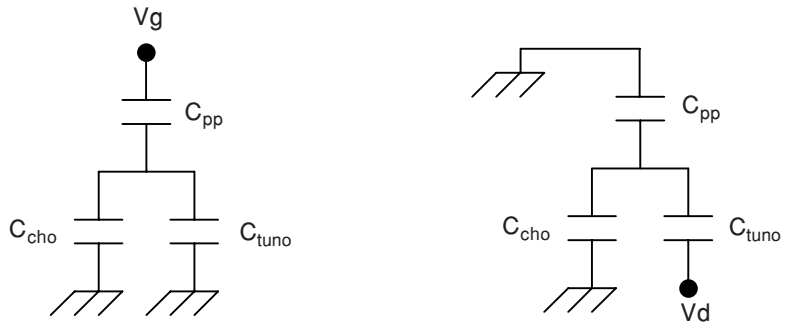
Figure 2 Memory Cell of EEPROM Device

metal CMOS process with three different oxide thickness which are a thin tunnel oxide (~130 A) used in memory cells, a thin gate oxide (~200 A) used in the low voltage transistor and a thick gate oxide (~500 A) used in the high voltage transistor. The first poly layer is used as the floating gate and the second poly is used as the control gate.

EEPROM device in a logic process requires a very high quality tunnel and interpoly capacitors with ultra thin dielectrics giving a high capacitance per unit area and very low leakage currents. The combination of silicon dioxide (O) and silicon nitride (N) materials can achieve the performances required, especially if a ONO structure is adopted.

The ONO dielectric is a multi-layer of silicon oxide (O), nitride (N) and silicon oxide(O) between two doped polysilicon layers as shown in Figure 1. These different layers have to be typically 5-10 nm thick. Consequently conventional measurements techniques such as ellipsometry or interferometry cannot be used to characterize them. We had to implement physical measurements, such as FESEM or FIB, to characterize accurately the contribution of each layer in the global ONO multi-layer.

Figure 3 shows the simplified equivalent circuit of the EEPROM transistor during program high and program low operations respectively. C_{pp} is the poly to poly overlap capacitance, C_{tuno} is the tunnel window capacitance and C_{cho} is the channel capacitance [6].



Programming high (Write) operation Programming low (erase) operation

Figure 3 Simplified equivalent circuit of the FLOTOX during the program high and program low operations

The programming speed of the EEPROM cell depends critically on Tunnel Oxide Thickness (X_{tun}), Programming Voltage (V_p), ONO Thickness (X_{pp}) and Poly to Poly overlap Area (A_{pp}). However, in this experiment only ONO and tunnel oxide layer are optimized and characterized.

The thinner the tunnel oxide is, the faster the programming speed. However, due to the charge retention requirement, the tunnel oxide thickness cannot be reduced beyond 85Å. Therefore 85Å is the nominal thickness throughout the experiment and will not be changed.

2.0 EXPERIMENT METHODOLOGY

This section of the paper will describe the methods of carrying out the experiment. This includes fabrication of the Floating Gate Tunnel Oxide EEPROM cell and experiment for the improvement of the programming speed. The patterning step, which consists of photolithography and etching process, are not described in detail in this paper.

A. Fabrication of the EEPROM cell

The EEPROM memory cells under study are fabricated on p-type silicon wafers <100> with the resistivities of 16–25 ohm-cm. The fabrication of the memory cell started with implantation of arsenic to form the buried N^+ region. Thermal oxide was subsequently blanket grown to a 600Å thickness on the buried N^+ region. The tunnel window was then patterned where ultra thin tunnel oxide was immediately grown to a thickness of 85Å. The formation of the floating gate was then carried out by the deposition of $POCL_3$ doped polysilicon with a thickness of 2300Å. ONO was used for inter-poly dielectric to insulate the floating gate from other electrodes. The ONO dielectric consists of a 100Å thermal oxide layer grown at 980°C in an O_2/N_2 environment, an LPCVD nitride layer of 230Å thickness deposited on top by CVD furnace and lastly a thermal oxide layer of 20Å thickness. The inter-poly dielectric construction was accomplished with the patterning of ONO. Next the formation of the control gate was done by deposition of polysilicon of 4000Å thickness using LPCVD. The fabrication process was continued with conventional backend processes for the interconnect construction to complete the whole circuitry. The n-channel memory cell transistor structures are now ready to be optimized for programming speed improvement. Focus ion beam was employed in this experiment to observe the cross section view of ONO and tunnel oxide layer of the EEPROM cell.

B. ONO Thickness Optimization

The first step to improve the programming speed is by scaling down the ONO layer. In this experiment only nitride layer is scaled down whereby the top and bottom oxide were fixed at 100Å and 20Å respectively. The nitride layer was deposited at three different thicknesses ie. 100A, 130A and 160A for the formation of ONO layer of EEPROM cell using standard procedure as mentioned in section 2A.

The Agilent 4155C DC Parameter Analyser was used to measure the threshold voltage and breakdown voltage at three different nitride thicknesses. From the breakdown voltage measurement results, the capacitor that exhibits greater than 1uA current is regarded as the breakdown of that capacitor.

The setup for the measurement of the threshold voltage to characterize the programming speed is as follows. An Agilent 4155C DC Parametric Analyser was

used to generate the "high" and "low" waveform signals for the programming operation. First, the EEPROM cells were exposed to a 5ms program low pulse of 16V to bring the Vt to a low level. The cell was then programmed high with a fixed programming voltage but with the variable programming pulse width. The programming pulses increased so that the cumulative programming time falls in logarithmic increments of 0.5ms, 1ms, 2ms, 5ms, 10ms ...100ms. Threshold voltage was measured at each of these cumulative intervals. After a complete sweep of exposure times up to 100ms, the programming voltage was incremented and the process was repeated again. Similar procedures were used to measure the program down speed by programming the cell high at the start and then programming low successively.

C. Oxide-Nitride-Oxide Integrity

The ONO integrity check was carried out using a new wafer. For this reason, capacitors were fabricated on n-type silicon wafers (100). A uniform layer of poly1 was then deposited across the wafer followed by the ONO layer formation consisting of bottom oxide, nitride and top oxide. The thickness of ONO layer is 100, 130 and 20 angstroms for bottom oxide, nitride and top oxide respectively. Finally a second polysilicon layer (poly2) was patterned on top of ONO layer to form 'large' capacitors 0.0308cm^2 in size only. The wafer was subsequently exposed to backside reactive ion etching to remove any oxide to facilitate backside substrate contact via a chuck. The final structure of the capacitor for ONO integrity check is shown in Figure 4. Breakdown voltages were then measured and recorded in histogram format. Separate sweeps were taken consisting of those with the substrate grounded while the polysilicon 2 pad is positive and those with the substrate is grounded while the polysilicon 2 pad is negative.

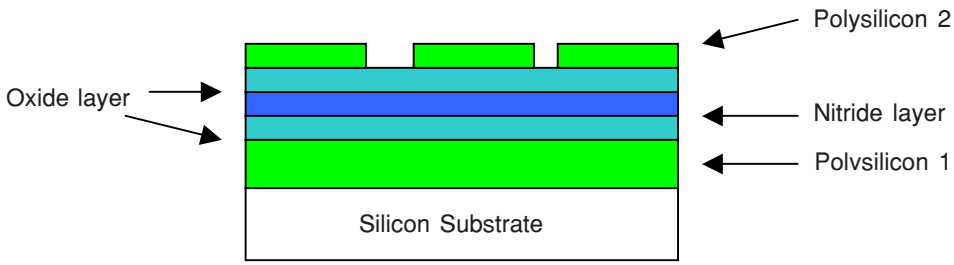


Figure 4 Capacitor structure for ONO layer integrity check

D. Tunnel Oxide Integrity

To facilitate tunnel oxide integrity test, capacitors of varying sizes were fabricated with the tunnel oxide as the dielectric. FN tunneling currents were then induced by applying

a ramp voltage across the capacitors. The breakdown voltage of each capacitor was then measured where the breakdown voltage is defined as the voltage applied across the capacitor to induce 1μA/cm^2 of tunneling current. The results were then plotted in a histogram.

For the tunnel oxide, capacitors were fabricated on p-type wafers (100). A uniform layer of tunnel oxide was deposited directly on the substrate using the relevant process steps. The thickness of tunnel oxide is 85 angstroms. Subsequently, a metal layer was patterned on top of tunnel oxide to form 'small' capacitors, which is 100 μm by 100μm in size (the size of a standard bond pad) and a 'large' die sized capacitors, which is 0.0308 cm^2. As of the ONO wafers, these wafers were also exposed to backside reactive ion etching to enable backside wafer contact via a chuck. Because of the p-type substrate, positive voltage was always applied to the substrate. Breakdown voltages were then measured and recorded in histogram format. The final structure of the capacitor for Tunnel Oxide integrity check is shown in Figure 5.

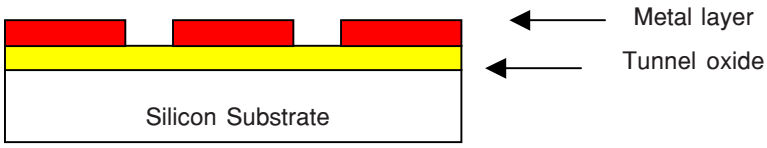


Figure 5 Capacitor structure for tunnel oxide integrity check

3.0 RESULTS AND DISCUSSIONS

Figure 6 is the focus ion beam cross-sectional view of the EEPROM cell. This figure clearly shows a complete Floating Gate Tunnel Oxide EEPROM cell, which consists

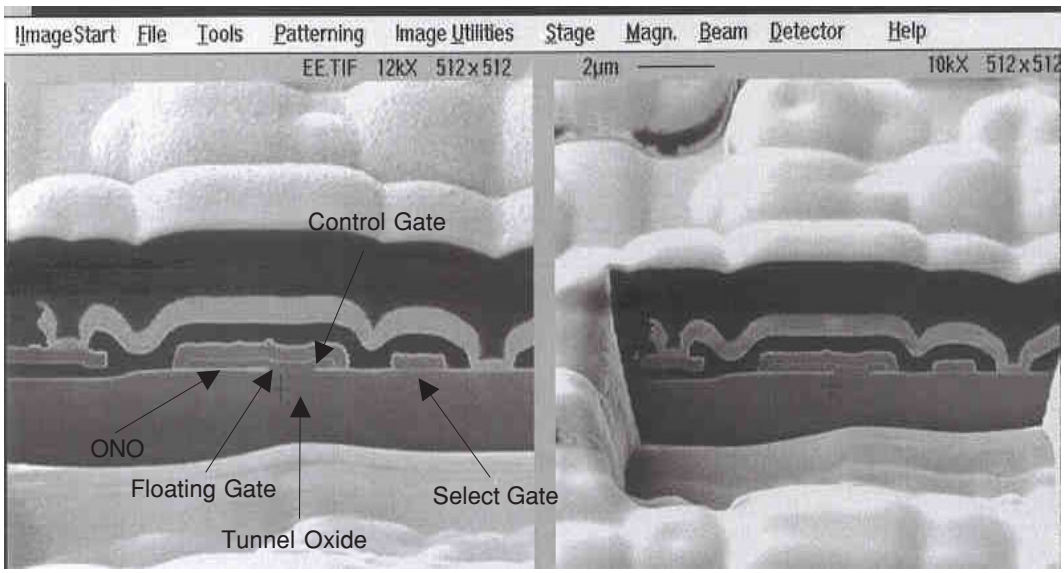


Figure 6 FIB cross-sectional view of a EEPROM cell

of a control gate, a floating gate, a select gate, a tunnel window, an ONO layer, a source area and a drain area. The EEPROM cell is sitting on a buried N^+ layer. The size of tunnel oxide window is approximately 1.0 um × 1.0 um on silicon and the ONO layer of approximate 130 angstroms is sandwich by two-polysilicon gate.

The programming performance of the EEPROM cell mostly depends on the thickness of the ONO layer. Therefore, the original thickness of the ONO layer was scaled down to improve the programming performance. As been discussed earlier, ONO is a sandwich layer of oxide-nitride-oxide. In this experiment, only the nitride layer was scaled down. Before scaling down, the original nitride thickness in the ONO layer was 230Å.

After scaling down, the final optimized thickness of the ONO layer must be able to withstand 16V for reliability purpose. Figure 7 illustrates the effect of different nitride thickness in the ONO layer to Vt-high and Vt-low whereby Figure 8 illustrates the effect of different nitride thickness in the ONO layer to the ONO breakdown voltages.

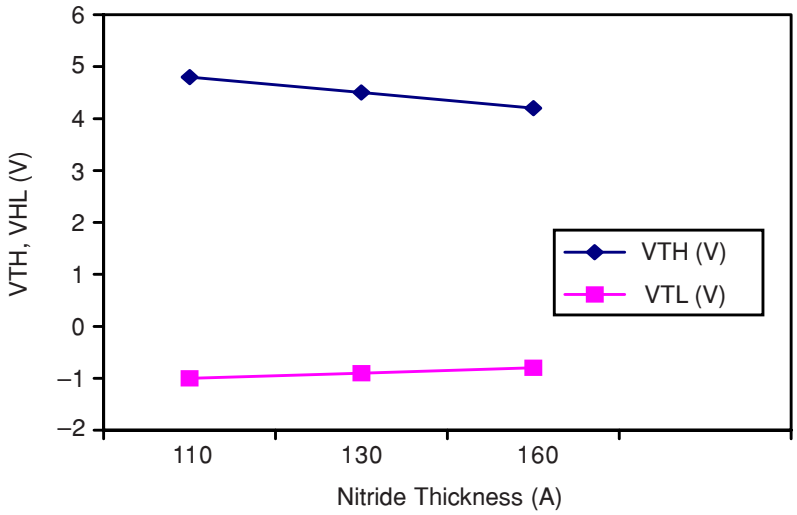


Figure 7 The high and low cell threshold voltage versus the nitride thickness in the ONO layer

The experiment on the programming speed versus ONO thickness shows that thinner ONO gives faster programming speed. Figure 7 clearly shows that the window of the Vt-high and Vt-low is narrowing down with the incremental of nitride thickness. The Vt-low is increased by 0.2V and Vt-high is decreased by 0.6V from 110Å to 160Å of nitride thickness. On the other hand, Figure 8 clearly shows that ONO layer breakdown is increased almost linearly with nitride thickness of ONO layer. It is increased by 1.3V and 2.3V of ONO layer breakdown voltage for 130Å and 160Å, respectively, from initial nitride thickness of 110Å. However, the specification of ONO breakdown voltage is set at 16V. In such a case, 130Å is chosen to be the standard for nitride thickness. With this thickness VtH reaches 4.6V and VtL reaches –0.9V in

1ms. This is enough for successful EEPROM operation. From Figure 8, the breakdown voltage of ONO layer at 130Å is 16.3V. This value is still above the specification set which is 16V.

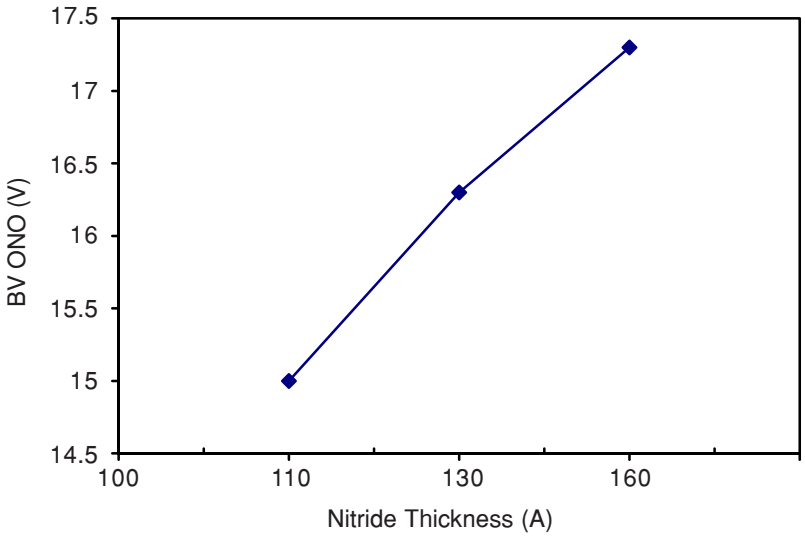


Figure 8 ONO breakdown voltage versus nitride thickness in the ONO layer

Figure 9 and Figure 10 are two histograms showing the ONO voltage distribution. The first histogram, which is Figure 9, was taken with poly2 undergoing a positive voltage sweep and the substrate grounded. The second plot, which is Figure 10, had

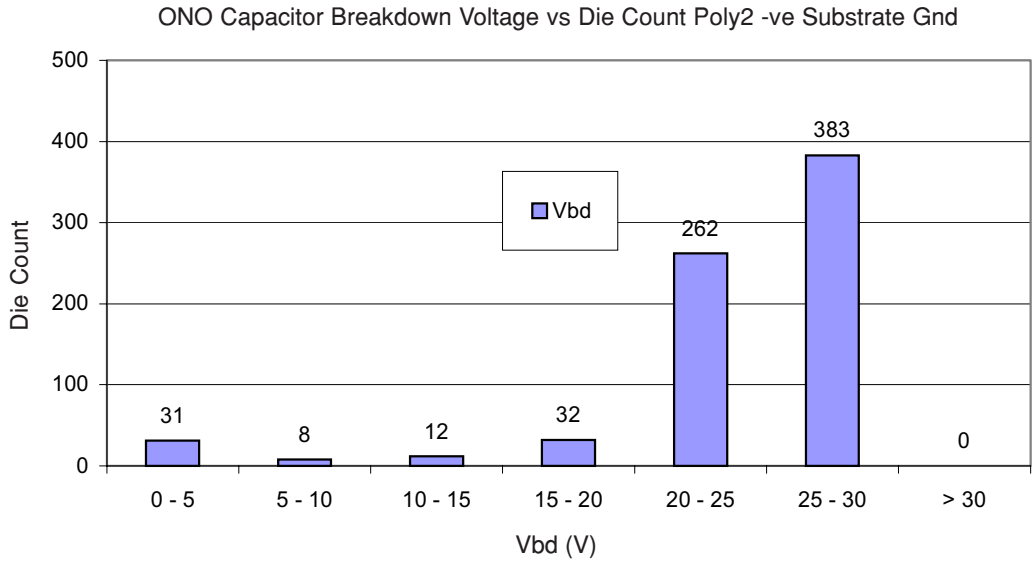


Figure 9 ONO Capacitor Breakdown voltage distribution with poly2 swept +ve and substrate grounded

poly2 undergo a negative voltage sweep with the substrate grounded. Using equation (2), the calculated defect density for both plots are about the same, 88.5% and 86.1% for Figure 9 and Figure 10 respectively. It shows that positive and negative voltage sweep while the substrate is grounded does not give any significant impact to the defect density of the ONO Layer.

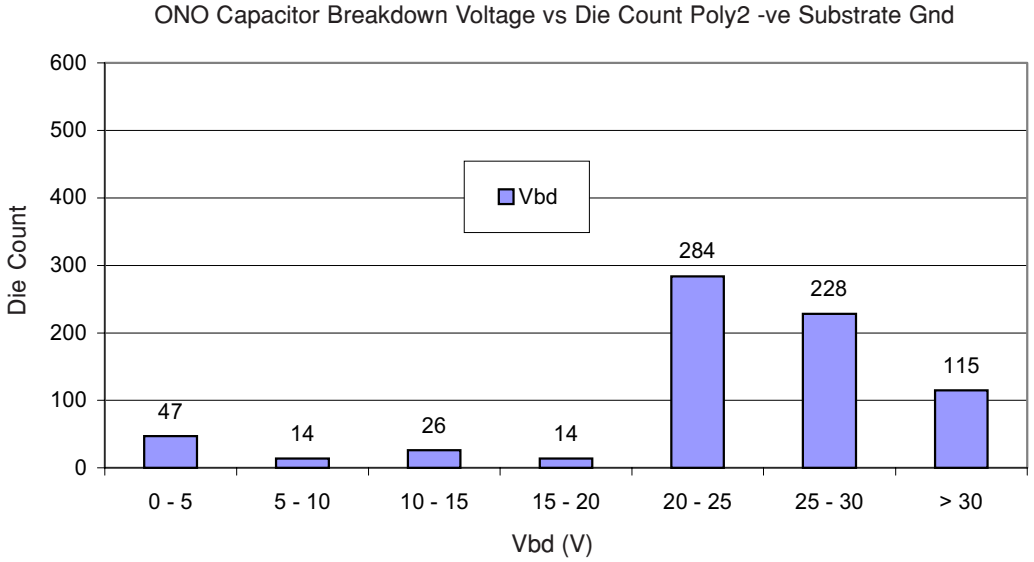


Figure 10 ONO Capacitor Breakdown voltage distribution with poly2 swept -ve and the substrate grounded

Yield is given by the equation below:

$$Y = \frac{1}{1 + AD} \tag{1}$$

where Y = yield
A = area
D = defect density

The capacitor area fabricated in this experiment was 0.0308 cm^2. The actual area used in a 16 kbit chip is substantially smaller at 0.00149 cm^2 and giving a step yield of 98.7% which is satisfactory.

Figure 11 shows the current density versus electric field (JE) relationship of the tunnel oxide. Site 1 (dotted line) is a defective site shown here with a good site (site 2) for comparison. We can see that the two sites possess the same characteristics up till 13.9 million V/cm where the defective site has visibly higher leakage.

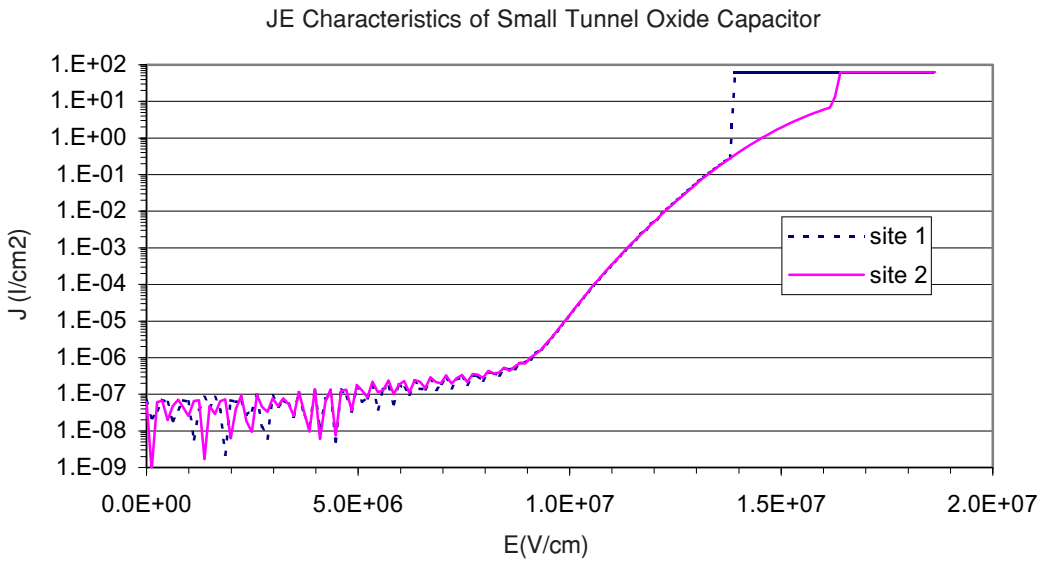


Figure 11 JE characteristics of a defective site (site1) versus a good site (site2)

The next two plots are histograms showing the distribution of the breakdown voltages for the small bondpad-sized capacitors (Figure 12) and the large die-sized capacitors (Figure 13). The thing to note about these results is that for the bulk of the sites tested, the breakdown is higher for the small capacitors (>10V) than the large capacitors (8–10V). This can be attributed to the fact that the large capacitors are more likely to pick up defects such as dust particles due to their size resulting in lower breakdown voltage distribution. Also there are much less defective sites that breakdown between 0–8V for the small capacitors.

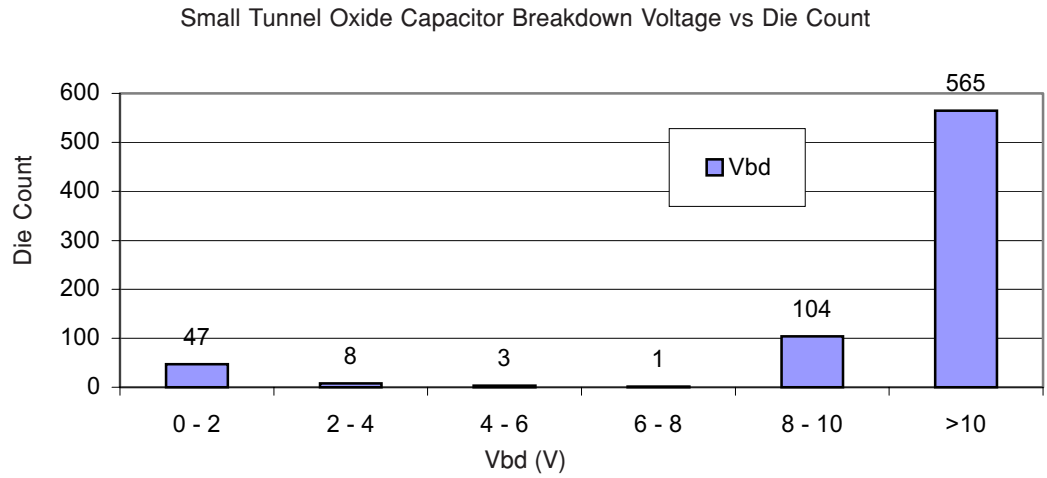


Figure 12 Small capacitor breakdown voltage distribution

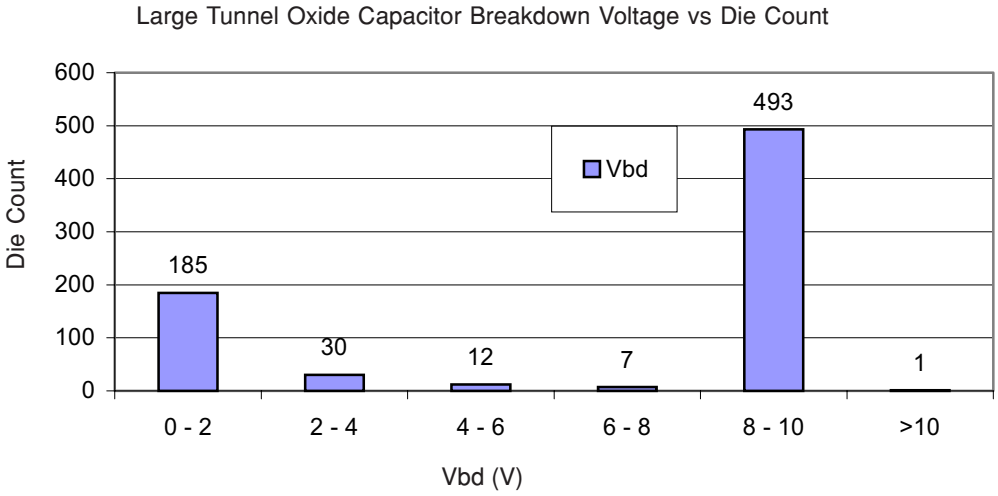


Figure 13 Large capacitor breakdown voltage distribution

The passing Vbd for the capacitors is defined as 8V and above. The large capacitors have an area of 0.0308 cm^2. This gives a defect density of 11.002 defects/cm^2 and a yield of 55.9%. However for a standard 16 kbit EERPOM chip the area of tunnelling oxide required is 0.78 μm^2. With this area size, the yield becomes a satisfactory 99.92%.

From the above result it shows that the current facility and process specification setup able to produce memory devices with high speed at high yield.

4.0 CONCLUSION

The characterization work to increase the programming speed of the memory cell of 16k EEPROM has been carried out. After scaling down the nitride of ONO layer from 160Å to 130Å, the Vt program window are further improved from 4.3V to 4.5V and from –0.7V to –0.9V for program high and low operations respectively. As the specification of the ONO breakdown voltage is set at 16V, 130Å was found to be the best thickness for nitride of ONO layer. The breakdown voltage for ONO at 130Å of nitride thickness is 16.3V. The experiment revealed that the yields of ONO and tunnel oxide layer of the actual size on silicon were achieved at 98.7% and 99.92%, respectively.

ACKNOWLEDGEMENT

The authors would like to thank MIMOS Berhad FAB1, Failure Analysis lab and Test lab personnel for their technical advice and contributions.

REFERENCE

[1] William, D. B. and J. E. Brewer. Nonvolatile Semiconductor Memory Technology, A Comprehensive Guide to Understanding and Using NVSM Devices. *IEEE Press*. 1-149.

[2] Canet, P., R. Bouchakour and J. M. Mirabel. 2001. EEPROM Programming Study – Time and Degradation Aspects. *IEEE letter*. 846-849.

[3] Robert, V., Augusto Pieracci and Bruno Ricco. 2001. Fast Programming/Erasing of Thin-Oxide EEPROMs. *IEEE Transactions on Electron Devices*. 48(4): 817-820.

[4] Dennis Walker, Jr. Brad Skotko and Steven Bibyk. 2001. Charactericzation of Two Standard CMOS EEPROM Designs. *IEEE letter*. 894-896.

[5] Bhattacharyya, A. 1984. Modelling of Write/Erase and Charge Retention Characteristics of Floating Gate EEPROM Devices. *Solid State Electrons*. 27(10): 899-906.

[6] Kolodny, A., S. T. K. Neih and B. Eithan. 1986. Analysis and modeling of Floating Gate EEPROM Cell. *IEEE Trans. Electron Device*. ED33(6): 835-844.

[7] Yaron, G., S. J. Prasad, M. S. Ebel and B. M. K. Leong. 1982. A 16k EEPROM Employing New Array Architecture and Design-in Reliability Features. *IEEE J. Solid State Circuits*. SC-17(5): 833-840.