14th Annual Microelectronic Engineering Conference, 1996.

Flash EEPROM Fabrication

Keith E. Zawadzki
Microelectronic Engineering
Rochester Institute of Technology

Abstract- Rochester Institute of Technology's (RIT's) first flash EEPROM cell was designed, fabricated and tested at its Microelectronic Engineering Fabrication Facilities. The RIT flash memory cell was modeled after Intel's flash EEPROM cell and realized a channel length = 8μm and channel width = 32μm. The fabrication technology required critical process development steps including a high quality 100Å tunnel oxide, 100 Å polySi oxide and a 200Å reactive sputter-deposited silicon nitride film. The cell successfully demonstrated 10 program and erase cycles and a programmed charge state for 130 hours before unacceptable charge loss occurred.

I. Introduction

Flash EEPROMs (Electrically Erasable and Programmable Read Only Memory) have become an attractive solution for fast (flash) erasable and programmable non-volatile memories, replacing the standard UV light erase PROM. The growing target market for flash technology is mass storage for mobile computing. This is a result of flash technology's many attributes, including; physically small, light weight, low power, high performance, reliable and mechanically rugged against vibration and shock. Current flash EEPROMs are available in 16Mbit sizes while the memory industry races to market 64Mbit flash.

The typical Intel, AMD and Sharp flash EEPROM cell is a basic NMOSFET with the exception of an additional capacitor. This additional capacitor is formed by a polySi control gate ontop of an oxide-nitride-oxide, or *ONO*, dielectric stack (see Figure 1). Note that the polySi *floating* gate is completely isolated from (1) the NMOS channel by a thin tunnel oxide, and (2) the polySi control gate by the ONO dielectric layer. Thus, the flash EEPROM cell is formed by two capacitors in series, whereas the top plate is formed by the polySi control gate, the middle plate is an isolated or *floating* ploySi gate, and the bottom plate is the NMOS channel of the P-well substrate.

The study of current semiconductor industry process technology in academia is crucial for preparing young engineers for the ever changing and competitive job market. This effort is present at Rochester Institute of Technology, whereas the Microelectronic Engineering program strives to prepare undergraduate students with today's IC technologies. RIT's Factory Pwell CMOS process utilizes a RIT testchip consisting of six different dimension NMOSFETs among many other test devices. Process integration was implemented to modify RIT's Factory Pwell CMOS process to create individual flash EEPROM cells from the various dimension NMOSFETs (see Figure 2). The anticipated result was to create a single flash EEPROM cell, similar to the Intel flash cell, and demonstrate the ability to electrically program and then electrically erase the cell. The definition of program was to increase the threshold voltage (Vt) of the cell by > 1V. The definition of erase was to then decrease the programmed cell's Vt by > 0.8 V.

II. Process Technology

RIT's Factory Pwell CMOS process is a one metal layer process which requires nine mask levels and sixty process steps. Critical changes within the sixty steps were required to fabricate a flash EEPROM cell from the NMOSFET. VLSI design changes included redesigning the contact cut and metal mask levels. The most critical and new process changes within RIT, included (1) growing a high quality 100Å tunnel oxide, (2) thermally oxidizing the polySi floating gate to create a 100Å polySi oxide and (3) reactive sputter-depositing a 200Å silicon nitride film on top of

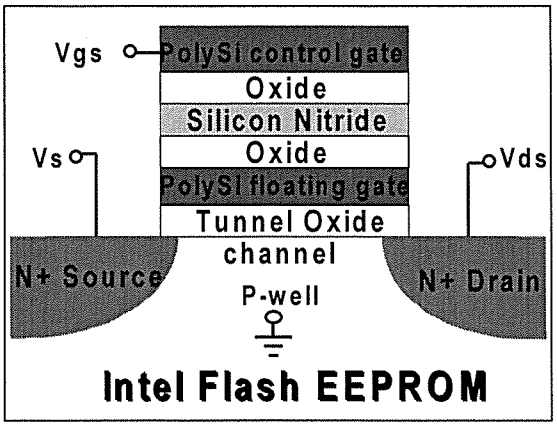

Figure 1: Intel Flash EEPROM cell

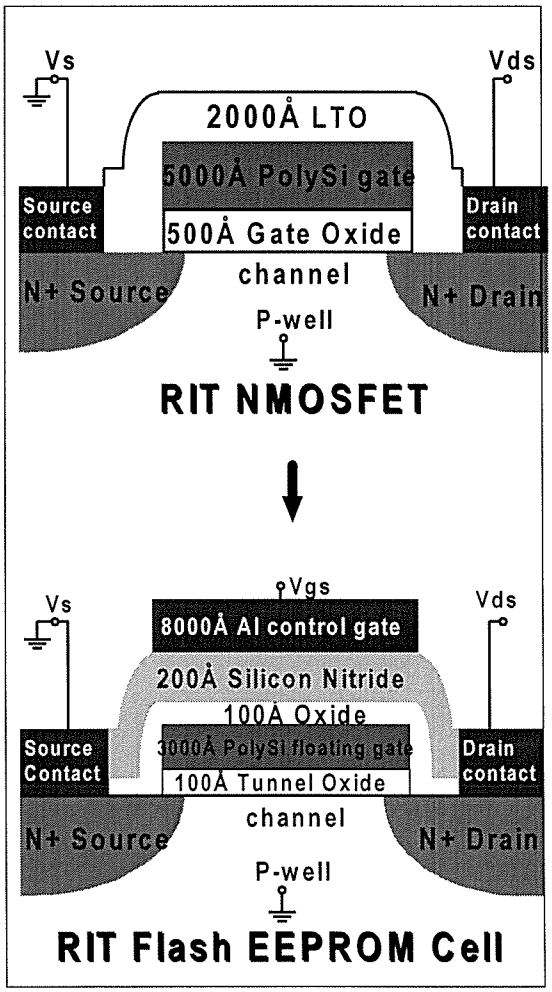

Figure 2: Project Goal- To modify RIT's Factory Pwell CMOS process to transform the RIT NMOSFET into the RIT Flash EEPROM cell.

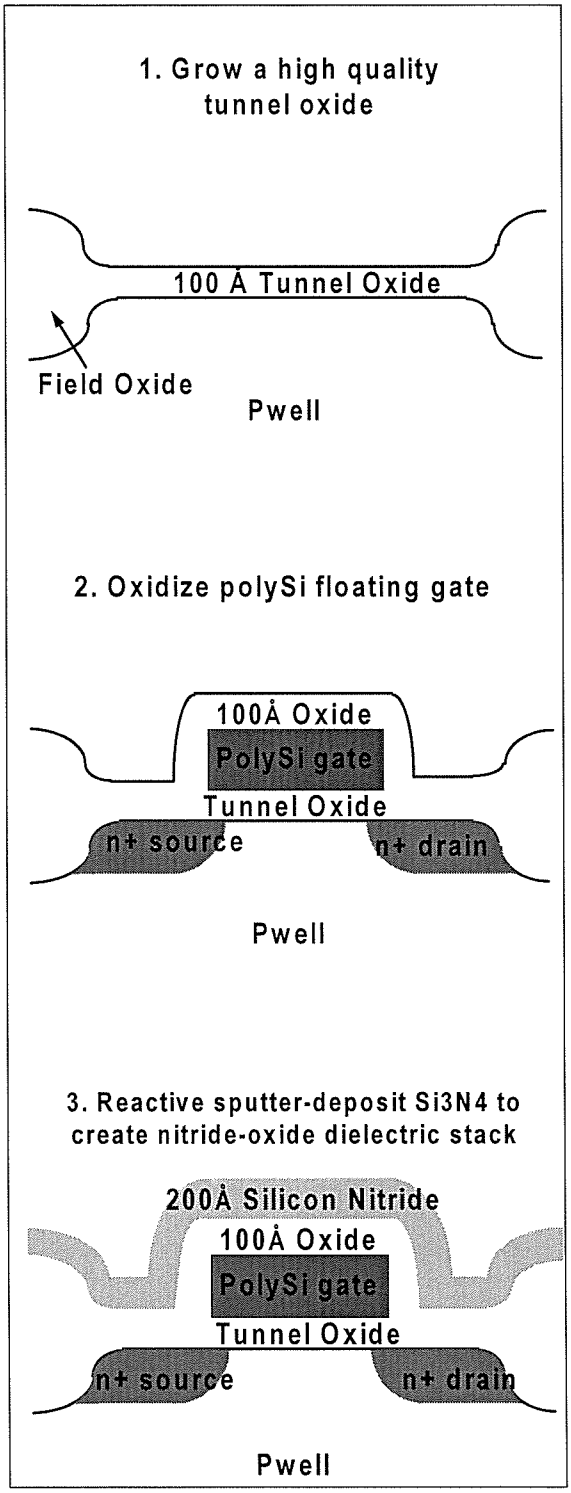

Figure 3: The three most critical and new process changes to create RIT's flash memory cell.

the polySi oxide to realize a nitride-oxide dielectric stack. These three critical process changes are illustrated in Figure 3. The integrity of all three of these films was essential to prevent any leakage paths which would destroy the flash EEPROM cell operation.

The 100Å tunnel oxide recipe was modeled after a former RIT MS student thesis on thin oxide leakage [Ref. 1]. This 100Å tunnel oxide process recipe using a Thermco furnace is listed in Figure 4.

After the tunnel oxide growth, 3000Å polySi was immediately low pressure chemically vapor deposited (LPCVD) ontop of the tunnel ox to cap the oxide and avoid contamination. The polySi was then patterned and etched to form polySi floating gates. After performing the necessary N+ source/drain lithography, the N+ source/drain implant followed. This N+ S/D implant was also used as the polySi dopant implant. This doping of the polySi floating gate added complexity to the 2nd critical process change, the polySi oxidation step. The oxidation rate of polySi is greater than crystalline Si and is a strong function of dopant in silicon. Thus, the N+ polySi dopant must first be annealed to provide uniform doping across the polySi floating gates and then oxidized. Furthermore, it is essential not to drive the N+ polySi dopant through the tunnel oxide. Diffusion through the tunnel oxide could degrade the tunnel oxide integrity and lead to a leaky tunnel ox. Therefore, the 100Å tunnel oxide process was modified for the 100Å polySi oxide process. The major reason for modification was a thermal budget to avoid driving the implanted polySi dopant through the tunnel oxide. The 100Å dry polySi oxide recipe using a Thermco furnace is shown in Figure 5.

The third critical process step was to reactive sputter-deposit 200Å silicon nitride ontop of the polySi oxide to form a nitride-oxide dielectric stack. The Perkin-Elmer 2400 Sputter system was used to reactive sputter-deposit single wafers at a time. A silicon target was installed and sputtering took place in a argon and nitrogen ambient at chamber pressure = 7mTorr. Because sputtering does not take place inside a perfect vacuum, longer sputtering times means an increased rate of oxidation due to some fab air inside the chamber. Longer sputter times could lead to a silicon-oxynitride film instead of a silicon nitride film. As a result, the sputter time was set to a relatively short time of 5 minutes. The only factor used to target the 200Å silicon nitride film thickness was the power. A short experiment was performed to characterize the PE 2400 sputter system to show thickness vs. power for the given set-up parameters. Furthermore, to densify the sputtered film and reduce pinholes, the film was annealed in a nitrogen ambient for 10 min at 850C inside a Thermco furnace. This experiment, including the results, are included in Figure 6. As illustrated, setting the power = 300W will result in a film thickness of 205.5 Å after the anneal step. It is also significant to note the index of refraction for the sputtered films. The index of refraction (n_i) for silicon nitride is 2.0. Thus, sputtering at a lower power setting yielded higher n_i values, meaning silicon nitride quality was sacrificed as power was reduced. After annealing, n_i for all the films improved to approximately $n_i = 2.0$. Thus, the annealing step in N2 may have had some chemical ramifications in providing an improved silicon nitride film.

III. Results and Discussion

The result was a successful RIT flash EEPROM cell which was both electrically programmable and electrically erasable. The cell (see Figure 2) tested had a channel length = 8μm and channel width = 32μm. The Id vs. Vgs curves for the cell in its initial, program and erase state are shown in Figure 7. The program and erase pulse train operated at 10,000 cycles/second (10kHz) and 14V amplitude. To create a 1.725V threshold voltage (Vt) shift as featured in Figure 7, the program operation required the gate to be pulsed for 30,000 cycles (3 seconds) while the erase operation required the drain to be pulsed for 40,000 cycles (4 seconds). Programming the cell by shifting the Vt from 0V to 1.725V was accomplished by Fowler-Nordheim tunneling electrons through the tunnel oxide and onto the polySi floating gate. Erasing the cell, by shifting the Vt from 1.725V back down to 0V, was performed by Fowler-Nordheim tunneling the electrons off the polySi floating gate, through the tunnel oxide and into the drain. Thus, programming stored electrons on the polySi floating gate, while electrons removed the stored electrons from the polySi floating gate.

RIT 100Å TUNNEL OXIDE PROCESS				
step#	tool: Thermco Furnace	tube	gas flows (lpm)	
	step description	temp (C)	N2	dry O2
1	ramp tube to 650C	650	8.5	0
2	push wafers in at 650C	650	8.5	0
3	10 min stabilization time	650	8.5	0
4	21 min tube ramp to 950C	650 to 950	8.5	0
5	16 min first oxidation	950	8.5	2
6	30 min first N2 anneal and ramp	950 to1050	8.5	0
7	25 min ramp down to 950C	1050 to 950	8.5	0
8	16 min second oxidation	950	8.5	2
9	30 min second N2 anneal	950	8.5	0
10	1.5 hr tube ramp down	950 to 650	8.5	0
11	pull wafers at 650C	650	8.5	0

Figure 4: RIT 100Å Tunnel Oxidation Process

RIT 100Å POLY-Si OXIDATION PROCESS				
step#	tool: Thermco Furnace	tube	gas flows (lpm)	
	step description	temp (C)	N2	dry O2
1	ramp tube to 650C	650	8.5	0
2	push wafers in at 650C	650	8.5	0
3	21 min tube ramp to 950C	650 to 950	8.5	0
4	10 min N2 anneal	950	8.5	0
5	15 min first oxidation	950	8.5	2
6	13 min first N2 anneal	950	8.5	0
7	15 min second oxidation	950	8.5	2
8	1.5 hr tube ramp down and anneal	950 to 650	8.5	0
9	pull wafers at 650C	650	8.5	0

Figure 5: RIT 100Å Poly-Si Oxidation Process

RIT Reactive Sputter-Deposit 200Å Silicon Nitride Process	
tool:	Perkin Elmer 2400 Sputter System
mode of operation:	single wafer sputter-deposit
target:	Silicon
Pumpdown Pressure:	6.7e-7 Torr
Gas Flows:	Argon 15sccm
	Nitrogen 15sccm
Chamber Pressure:	7mTorr (in process)
Time:	5min
Power:	300W
Anneal step:	Thermco Furnace
	temp: 850C
	gas flow: N2 5lpm
	time: 10min

Silicon Nitride Reactive Sputter-Deposition Experiment to Target 200Å film thickness (same tool set-up as described above, except for POWER variable)				
Power (W)	initial film		after anneal	
	thickness (Å)	Index of Refr	thickness (Å)	Index of Refr
100	75.9	1.7	71.9	2.05
200	176	2.5	158	1.95
300	208	2.3	205.5	1.95
400	292	2.1	286	2

Figure 6: RIT Reactive Sputter-Deposit 200Å Silicon Nitride Process and Experiment to target Power vs. Thickness.

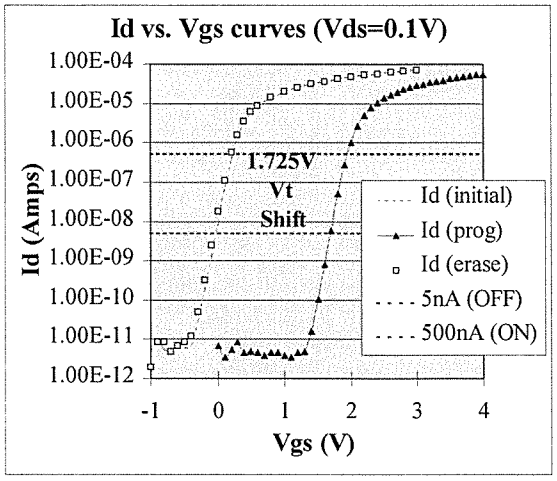

Figure 7: Id vs. Vgs curves for the RIT Flash memory cell in its initial, programmed and erase state. [Assume RIT read operation Vgs = 1V]

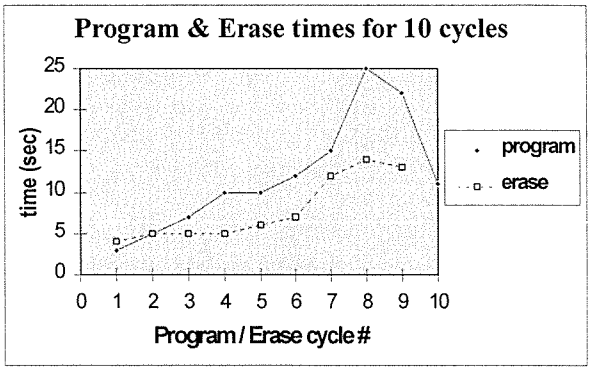

Figure 8: Program and erase pulse times for each successive P/E cycle. [NOTE- Program and erase time increases for each cycle, displaying non-ideal flash preformance]

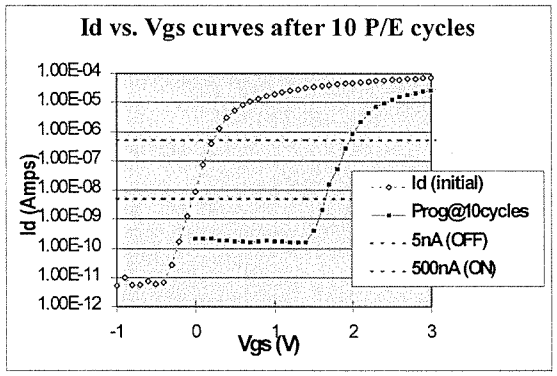

Figure 9: Id vs. Vgs curves after 10 P/E cycles to illustrate increase in leakage current, implying tunnel oxide degradation.

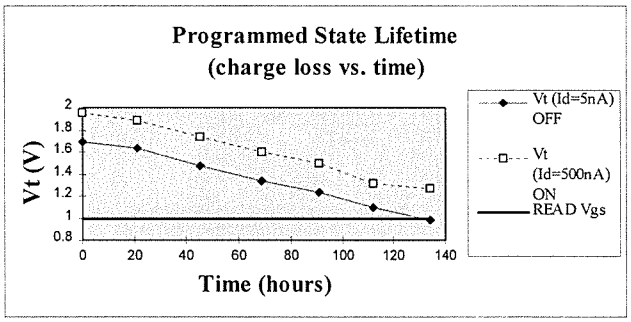

Figure 10: Programmed state lifetime to illustrate that the RIT flash cell is a non-volatile memory for up to 130 hours. [At time = 130 hours, the Vgs=1V read operation would fail to read a programmed state because the cell starts to turn-on]

Returning to Figure 7, if the cell current was > 500nA, the cell was ON, a binary "1". If the cell current was < 5nA, the cell was OFF, a binary "0". The Id vs. Vgs curves shows Vt_{erase_state} = 0V and $Vt_{program\_state}$= 1.725V. Thus, if Vgs=1V is used for a read operation, then the erased cell is ON during a read and the programmed cell is OFF during the read operation. This is the theory behind storing a binary "1" or a binary "0" in a flash EEPROM cell. If the cell is in its erased state, a "1" is stored. Whereas, if the cell is in a programmed state, a "0" is stored. Remembering the earlier discussion on programming and erasing, the difference between a stored "0" and a stored "1" in the flash cell is whether or not electrons are stored on the polySi floating gate ("0", program state) or are not stored on the polySi floating gate ("1", erase state).

The quality and reliability of the RIT flash EEPROM cell was researched. Figure 8 illustrates the program and erase time for successive program and erase (P/E) cycles. Figure 8 shows the cell required longer program and erase times for each successive cycle. Although this is a non-ideal flash cell characteristic, Figure 8 proves that the cell could successfully perform for at least 10 program and erase cycles.

A further look into the flash cell's performance after successive cycles is demonstrated in Figure 9. Figure 9 is the Id vs. Vgs curves replotted after 10 program and erase cycles compared to the initial curve. The significance of Figure 9 is that the leakage current of the cell has increased after 10 cycles from <10 pA range to 200pA. This again is a non-ideal flash characteristic implying tunnel oxide degradation due to electrons tunneling through and damaging the tunnel oxide. However, Figure 9 again proves the RIT flash cell will successfully perform for at least 10 P/E cycles.

Finally, to prove that the RIT flash EEPROM cell is a non-volatile memory, which is the essence of flash technology, the programmed state lifetime in shown in Figure 10. Figure 10 plots the $Vt_{program\ state}$ vs. time in hours. Vt is rapidly decreasing over time and at 130 hours after programming, the RIT flash cell would no longer correctly read a programmed state. Flash technology in industry is typically guaranteed for 10 years. However, although the RIT flash cell rapidly loses the stored electrons on the polySi floating gate compared to semiconductor industry standards, the cell was a successful non-volatile memory for 130 hours. Improvement to the non-volatile lifetime may be accomplished by depositing a passivation layer such as oxy-nitride to protect the wafer from contamination.

IV. Summary

Rochester Institute of Technology's 1st successful flash EEPROM cell was designed, fabricated and tested by applying process integration skills to modify RIT's Pwell CMOS process. The RIT flash memory cell realized a Vt shift of 1.725V by programming and erasing the cell. Quality and reliability testing illustrated 10 successful program and erase cycles and a non-volatile memory lifetime of 130 hours.

Acknowledgment

The author acknowledges Scott Blondell, Clay Reynolds and Dave Yackoff from RIT's Microelectronic Enginnering Fab for equipment and processing support and RIT faculty including Dr. Lynn Fuller, Dr. Santosh Kurinec and Dr. Dick Lane. The author also acknowledges Mr. Mark Webb, Intel Fab7 Device Manager for 0.4μm Flash Development Group.

References

[1] Rohan S. Braithwaite, "Thin Oxide Leakage" RIT MS Thesis, 1995.