Radiation and Reliability Concerns for Modern Nonvolatile Memory Technology


1Dell Services Federal Government, Inc., Fairfax, VA, 2MEI, Inc, Lanham, MD, 20706, 3NASA GSFC, Greenbelt, MD 20771, 4Naval Research Laboratory, Washington, DC 20375, 5Micro-RDC, Inc., Albuquerque, NM 87110

Abstract: Commercial nonvolatile memory technology is attractive for space applications, but radiation issues are serious concerns. In addition, we discuss combined radiation/reliability concerns which are only beginning to be addressed.

Introduction. Commercial nonvolatile memories are increasingly attractive for space applications because of their high level of integration, which means savings in size, weight, and power, and their extremely low cost per bit. In addition, the nonvolatile nature of these technologies makes them more resistant to some kinds of radiation effects than standard volatile memories. However, there are still some significant radiation concerns for these technologies, which include the effect of radiation exposure on their long term reliability. The dominant commercial technology is currently floating gate NAND flash memory, but concerns about its ability to support continued scaling have led the industry to investigate a variety of alternative technologies.

Description of the Technology. In floating gate flash memory, the storage element is a poly-Si gate which is completely surrounded by insulators. The cell is written by injecting electrons through a tunnel oxide, usually around 10 nm thick, at very high fields. The cell is erased by reversing the fields, and injecting the electrons through the oxide again, in the opposite direction. Both the program (write) operation and the erase operation require very high voltages, which are generated by an internal charge pump circuit. The charge pump often turns out to be the most sensitive part of the circuit for radiation damage. Normally, when functional failure occurs in a radiation test of a flash memory, it is because the part can no longer be erased, or written, or, frequently, both functions are lost at the same time. Usually, this means the charge pump can no longer put out the high voltages necessary to perform these operations.

In the NAND architecture, the bits are organized serially. For example, one source contact might serve for a string of 32 bits. In the alternative NOR architecture, the bits have random access, and each bit has its own contacts. For this reason, NAND is usually used for mass storage of data, but NOR is used to store operating instructions. Because NAND has fewer interconnects, it achieves higher bit densities, and, therefore, lower cost per bit. NAND typically achieves better total ionizing dose (TID) response than NOR, because the other
transistors in a string, which are normally off, tend to block any radiation-induced leakage current that might occur. In a NOR, there is nothing to block radiation-induced leakage current. For this reason, NORs typically fail from TID at about 10 krad (SiO₂), or at most a few tens of krad (SiO₂). NANDs, on the other hand, often survive 100 krad (SiO₂), which is adequate for most NASA space systems, although it still falls short of strategic military requirements [1,2]. In fact, the most recent TID test done in our group at GSFC was on the Samsung single chip, single level cell (SLC) 8G NAND, which survived past a dose of 400 krad (SiO₂). In a step-stress test, the first failures were at the next dose level, which was 500 krad (SiO₂).

For conventional volatile memories, cells upset when the voltage on a critical node is pulled down by an ion strike. But for NVM, including FG flash, the cells are designed to retain information with no voltage at all applied. For this reason, NVMs are typically several orders of magnitude less sensitive to SEU than standard volatile memories [3]. However, FG flash is typically a very complex circuit, with an on-chip processor to control the memory. Single event-induced errors in the control logic (SEFI—single event functional interrupt) are one of the two main concerns for SEE testing of FG flash. The other is destructive failure of the charge pump, which leaves the part unable to write, or to erase.

**Radiation and Reliability Concerns.** Besides simple radiation concerns, there is some concern that radiation might degrade the long-term reliability of FG flash. Radiation testing is usually done with fresh parts, but radiation effects at the end of life have not often been evaluated. It is well established in the literature that injecting charge through the oxide at high field in the Program (write) and Erase operations, causes damage to the oxide (for a review, see Mielke et al. [4]). One manifestation of this damage is SILC (Stress-Induced Leakage Current), which causes retention failures—a major reliability problem in flash memories, especially after repeated Program/Erase cycling. Charge leaking off the floating gate means the cells do not hold stored information as long as they are supposed to. The conventional model for SILC is that electrons tunnel from the floating gate to a trap, a stress-induced defect in the oxide, and then to another trap, and so on until the electrons reach the Si substrate. In a thin oxide, as few as two traps, properly aligned, can create a leakage path. Usually, the defects in the oxide that give rise to SILC are assumed to be trapped holes. It is also well established that radiation exposure can introduce defects, hole traps, which give rise to leakage current [5], which is called RILC (Radiation-Induced Leakage Current). Scarpa et al. [5] showed that hole trap-assisted tunneling was the underlying mechanism for both RILC and SILC. Since both RILC and SILC have the same underlying mechanism, one would expect that radiation exposure might have an effect on the reliability problems associated with SILC.

We note that, in [5], the authors reported measurable RILC, in the range of $10^{-11}$ A, but only after doses of several Mrad (SiO₂). Unhardened commercial technology would usually fail for other reasons, long before these doses were reached. On the other hand, the difference between a stored zero and a stored one is only about 67 electrons for a NAND flash cell with 60 nm technology, and less for newer technology. For ten year retention, which is the normal retention spec, leakage current has to
then be less than one electron every 54 days, which is about $3 \times 10^{-26}$ A, on average. Currents at this level are not directly measurable, but the question is whether or not they can be caused by the relatively low radiation doses that unhardened commercial technology can tolerate. In Fig. 1, we show results which suggest that radiation can cause end-of-life reliability failures in some cases.

Fig. 1. Error count for five samples of Samsung 8G NAND irradiated to 200 krad (SiO$_2$), compared to five unirradiated controls. Error counts varied from nearly 100 to almost 200 for irradiated samples, compared to one or two for controls.

The results in Fig. 1 were obtained by baking the irradiated samples and the controls at 100°C for over 1000 hours. The samples were not cycled, except for the minimal steps necessary to verify that the parts worked properly, and to store the initial pattern, which was a checkerboard. An important point here is that these parts were irradiated to 200 krad (SiO$_2$) which is enough to cause TID failure in most unhardened commercial parts. In Fig. 2, we show results from another test, but where the parts were irradiated only to 50 krad (SiO$_2$)—a much lower dose. In this case, there is some difference between the irradiated samples, and the unirradiated controls, but it is not statistically significant.

Fig. 2. (a) Error count for five irradiated Micron parts, compared to (b) unirradiated controls.

These Micron parts were irradiated, or not, to 50 krad (SiO$_2$), and then cycled to $10^5$ P/E cycles. Then their retention was monitored for 180 days (so far). Mean error count for the irradiated samples was 41 errors, compared with 24 errors for the controls. But the standard deviation for the two groups was 25 errors and 17 errors respectively. Since the difference between the groups was not larger than the variation within the groups, the difference cannot be considered statistically significant. However, one would expect the radiation effect to increase with dose, and at some larger dose, the difference would probably become significant, as it is for the
parts in Fig. 1. However, these parts failed for other reasons just slightly above 50 krads (SiO₂), so they would suffer TID failure before that dose could be reached.

An example illustrating retention failures after heavy ion exposure, from SEE, is shown in Fig. 3 [6]. In Fig. 3, the initial integral threshold voltage distribution before heavy ion exposure is shown (open circles), and also after exposure (open triangles). Cells hit by ions are shifted to the left, but when the part is reset, the initial distribution is essentially recovered (shaded diamonds). But after being reset, if the part is just allowed to sit, the cells damaged by the heavy ions start to leak charge, and a low Vt tail develops on the threshold voltage distribution. On the horizontal axis, the tic marks indicate intervals of 1V, and a shift of about 1V signifies a retention failure for that bit. Ions with LET less than about 30 MeV/mg/cm² do not cause this kind of retention failure—Xe ions, with LET about 58 were used to obtain the data in Fig. 3. There is strong experimental evidence that heavy ions with high LET can cause structural changes in SiO₂ [7].

Figure 3. Effect of heavy ion exposure, including retention failures, post-irradiation.

The most likely model for explaining the oxide structural modifications was proposed by Fleischer et al. [8]. They were cosmic ray physicists who used sheets of plastic or glass plates as cosmic ray detectors. They observed that insulator material around an ion track had a differential etch rate, which they attributed to structural modification. They considered a total of seven different models, and rejected all but what they called the “ion explosion spike” model. The basic idea was that the atoms along the track were so heavily ionized that coulomb repulsive forces between them caused the atoms to rearrange themselves, breaking chemical bonds. In this view, the structural modification is a coulomb effect, caused by intense ionization. They specifically rejected eight other models, including delta rays and displacement damage. The reason they rejected displacement damage from non-ionizing energy loss (NIEL) is that it would be concentrated at the end of the track, but the insulator structural modification takes place.
along the entire track—see, for example, Fig. 1 of [6] or the etching results of [8]. We note that this structural modification only happens in insulators, and not in semiconductors or metals, apparently because the greater density of mobile free carriers neutralizes the ionization along the track before the lattice can be disrupted.

Figure 4. Endurance errors for Micron 4G NAND flash, as a function of TID exposure.

However, it is clear that endurance failures are not accelerated by TID exposure, at least at the doses unhardened commercial technology can tolerate [9]. In Fig. 4, we show results where five parts were cycled at each dose level. The statistical variation within the groups at a given dose is greater than the variation with dose. Therefore, the conclusion was that the effect of TID exposure was not statistically significant.

Testing Issues and Lessons Learned. In testing of NVMs, and flash memory in particular, a number of practical testing issues have become apparent. First, angular effects are very important, both the difference between normal and high angle, and also between tilt and roll at a given angle. The difference between tilt and roll is the difference between having ions incident along the columns or across the columns, and results can be different. Destructive events in the charge pump happen primarily at normal incidence. To predict the failure rate in space, where the flux is omnidirectional, one has to have data at other angles.

Second, destructive failures happen primarily in the high voltage Program or Erase modes, which may have duty cycles of only 1-2%. In testing, one has to emphasize these modes, because these failures have the most severe consequences, but to predict the failure rate in space, the duty cycle correction has to be made.

Third, one has to try to keep the flux low enough to avoid collective effects. There have been reports [10, 11] claiming that there are high current spikes coming from the charge pumps of flash memories, which cause functional failures of the charge pumps. These spikes were said to be 80 mA or more, and 300-400 ms in duration. In an effort to replicate these results, we performed another experiment using test samples with the same part numbers, and the same beam conditions, but different test equipment. Although we observed 52 high current events, none matched the reported current spikes [12]. In Fig. 5, we show the current trace from one of the 38 beam runs, which has three high current events. In the entire experiment, 48 of the 52 high current events had a characteristic stair-step structure, where there was a change in the DC current level changed, or there was a series of changes, but the new current level persisted for anywhere from a few seconds to several minutes. The second event in Fig. 5 is one example of this type—current increases from about 5 mA to about 80 mA, and is stable for
about three minutes. In some of these events, the current will increase to a new value, and then, eventually, increase again, or recover spontaneously. These events are described by Shindou et al. [13] as LSEL (Localized Single Event Latchup), and are similar to what they observed in a combinational logic test chip.

In Fig. 6, the current is about 5 mA initially, which is the nominal Read current for this part, and which is appropriate, since the part is supposed to be reading. Then the current jumps to about 10 mA, which is the nominal Write current for this part. This suggests that some of the control logic is trying to Write, even though the commands being sent are to Read. After a few seconds the current increases rapidly, as if the Read and Write logic are fighting for control. This contention is resolved, when another Read command is received, and the part resumes Reading properly. Neither LSEL nor bus contention is unique to flash memory—both have been observed in anything containing combinational logic [13-15]. Indeed, the test vehicles in [13-15] do not even have charge pumps, so there is no obvious connection to the charge pump, even when there is one, as in flash memory. In order to identify the regions on the die that produce high currents, we did two other tests. The first of these used the NRL pulsed laser system, in a front surface, single photon absorption test. Using 590 nm (green) light, we obtained the results shown in Fig. 7.

The first high current event in Fig. 5 is one of the other four events, and appears to be similar to the bus contention, also described by Shindou [13]. This event is shown on an expanded time scale in Fig. 6.

Fig. 5. Current trace for Micron 4G NAND in Dynamic Read mode, irradiated with 2x10⁶ Xe ions/cm².

Fig. 6. Short transient current event from the beginning of Fig. 5, on an expanded time scale. High current lasts about 1 sec, baseline to baseline.

Fig. 7. Laser test results, where dark spots (red) indicate locations where high currents (>80 mA) were observed. Light spots indicate location of other SEFIs, but without high current.
In Fig. 7, the regions believed to be the charge pumps are indicated by white rectangles, and none of the high current spots are in the charge pump regions. Therefore, the connection between the charge pumps and the high currents is questionable, at best.

In the laser test we ran basically at full power, without any good way to convert that to an effective LET. We believe that running the laser at lower power would reduce, if not eliminate, the differences between the laser and the Milli-Beam. Further tests are planned to check this possibility. A more significant discrepancy is between the Milli-Beam and the broad beam results in [10,11] and other places. In [10,11], the authors had no difficulty producing numerous high current events with Xe ions, where the fluence for the whole die matched the Milli-Beam fluence for a small area. But in the broad beam exposures, the die was hit in perhaps 10,000 locations per second. The results under these conditions are qualitatively different than those when only one small region of the chip is hit at one time. In space, the particle flux is low enough that there will never be more than one part of the chip hit at given time. These results suggest that broad beam irradiations such as those in [10,11] are not useful for predicting the response of the parts in space.

Finally, to interpret test results, one has to be able to estimate the error rate in space from what is observed in the test. For example, at LET>60, the flux in space is about one ion/cm² per 125 years. Therefore, $10^4$ ions/cm² is equivalent to more than a million chip years in orbit, and $10^7$ ions/cm² is equivalent to more than a billion chip years in orbit. Therefore, effects observed only once in a million particles/cm² would, essentially, never happen in space.

Conclusions. Floating gate flash memories are attractive for space applications, but testing them is not straightforward. There are many issues to be aware of, and unknown unknowns to watch for.
Acknowledgments: The authors would like to thank Ken LaBel and the NASA NEPP program, and Bruce Wilson of DTRA for their support, and Martha O’Bryan for technical assistance with the conference presentation and this manuscript.

References.

12. T.R. Oldham et al., Heavy ion SEE test report for the current spike experiment,