National Aeronautics and Space Administration

Flash Memory Reliability
NEPP 2008 Task Final Report

Yuan Chen
Jet Propulsion Laboratory
Pasadena, California

Jet Propulsion Laboratory
California Institute of Technology
Pasadena, California

JPL Publication 09-9 3/09

National Aeronautics and Space Administration

Flash Memory Reliability
NEPP 2008 Task Final Report

NASA Electronic Parts and Packaging (NEPP) Program
Office of Safety and Mission Assurance

Yuan Chen
Jet Propulsion Laboratory
Pasadena, California

NASA WBS: 939904.01.11.10
JPL Project Number: 102197
Task Number: 1.23.6

Jet Propulsion Laboratory
4800 Oak Grove Drive
Pasadena, CA 91109

http://nepp.nasa.gov

This research was carried out at the Jet Propulsion Laboratory, California Institute of Technology, and was sponsored by the National Aeronautics and Space Administration Electronic Parts and Packaging (NEPP) Program.

Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise, does not constitute or imply its endorsement by the United States Government or the Jet Propulsion Laboratory, California Institute of Technology.

Copyright 2009. California Institute of Technology. Government sponsorship acknowledged.

Acknowledgment

The author would like to thank Integra Technologies for reliability testing support.

Table of Contents

Scope 1
Section 1. Flash Memory Technology 1
1.1 Memory Architecture: NOR versus NAND 2
1.2 Charge Storage: Floating Gate versus Discrete Storage 2
Section 2. Flash Memory Reliability 3
2.1 Bit Flipping 3
2.2 Bad-Block Handling 3
2.3 Endurance 4
2.4 Retention 4
2.5 Data-Loss Effects 4
Section 3. Reliability Testing on Samsung 8-Gbit NAND Multi-Level Cell (MLC) Flash Memory 5
3.1 Endurance 5
3.2 Retention 7
3.3 Summary 7
Section 4. Preliminary Thoughts on Risk Mitigation 8
4.1 Reliability 8
4.2 Redundancy 8
4.3 Media Management 9
References 10

Scope

This investigation is sponsored by the NASA Electronic Parts and Packaging (NEPP) Program. The objectives of this task are to evaluate the reliability of advanced flash memory technologies, to perform reliability testing on selected flash memory chips, and to provide recommendations for using the flash memory technologies in space applications from the perspective of both flash memory chip level and system level. This task leverages with another NEPP task, titled "Radiation Impacts on Flash Memory Technologies," to address the reliability and radiation effects on the same selected flash memory technologies and parts. This preliminary report is based on the flash memory reliability studies conducted during the first year (2008) of this multi-year task.

Section 1. Flash Memory Technology

Flash memory is nonvolatile memory that can be erased and reprogrammed in units of memory called blocks. In the mobile era, flash memory has been strongly driving the growth of the semiconductor memory business. In the near future, urgent needs of memory solutions in system-on–chip technologies and standard logic will boost the memory market even more. On the other hand, memory growth becomes increasingly difficult due to scaling limitations. Table 1 shows the scaling roadmap for flash memory technologies [1].

Table 1. Scaling Roadmap for Flash Memory Technologies.

Year of Production	*2005*	*2006*	*2007*	*2008*	*2009*	*2010*	*2011*	*2012*	*2013*
DRAM ½ Pitch (nm) (contacted)	*80*	*70*	*65*	*57*	*50*	*45*	*40*	*36*	*32*
Flash NOR L_g-stack (physical μm) [8], [9]	0.14	0.135	0.13	0.12	0.12	0.11	0.11	0.1	0.1
Flash NOR highest W/E voltage (V) [10], [11]	7-9	7-9	7-9	7-9	7-9	6-8	6-8	6-8	6-8
Flash NAND highest W/E voltage (V) [12]	17–19	17–19	15–17	15–17	15–17	15–17	15-17	15–17	15–17
Flash NOR I_{read} (μA) [13]	29–37	28–36	27–35	26–34	25–33	27–33	27–33	26–32	25–31
Flash coupling ratio [14]	0.65–0.75	0.6–0.7	0.6–0.7	0.6–0.7	0.6–0.7	0.6–0.7	0.6–0.7	0.6–0.7	0.6–0.7
Flash NOR tunnel oxide thickness EOT (nm) [15]	8–9	8–9	8–9	8–9	8–9	8	8	8	8
Flash NAND tunnel oxide thickness EOT (nm) [16]	7–8	7-8	6–7	6–7	6–7	6–7	6–7	6–7	6–7
Flash NOR interpoly dielectric thickness EOT (nm) [17]	13-15	13-15	13-15	13-15	13-15	10-12	10-12	10-12	10-12
Flash NAND interpoly dielectric thickness (nm) [18]	13–15	13–15	10–13	10–13	10–13	10–13	10–13	10–13	9–10
Flash endurance (erase/write cycles) [19]	1.00E+05	1.00E+05	1.00E+05	1.00E+05	1.00E+05	1.00E+06	1.00E+06	1.00E+06	1.00E+06
Flash nonvolatile data retention (years) [20]	10–20	10–20	10–20	10–20	10–20	10–20	10–20	10–20	20
Flash maximum number of bits per cell (MLC) [21]	2	2	2	2	2	4	4	4	4

1.1 Memory Architecture: NOR versus NAND

NOR and NAND technologies [2-4] dominate today's flash memory market. NOR flash memory devices, first introduced by Intel in 1988, revolutionized the market formerly dominated by Erasable Programmable Read-Only Memory (EPROM)- and Electrically Erasable Programmable Read-Only Memory (EEPROM)-based devices. The introduction of Toshiba's NAND flash architecture in 1989 addressed the need for lower cost per bit, higher-performance, and disk-like memory with a consistent interface for easy upgrade. In today's market, NOR dominates as a code-storage media and NAND for data storage. NAND has its strongest market presence in the memory card market (Compact Flash, Secure Digital, PC Cards, and MMC).

NOR is cost effective in low-capacity data storage, and delivers high read performance. A feature of NOR is eXecute In Place (XIP), which allows an application to be run directly from flash instead of reading the application code into system RAM. However, this usage suffers from extremely low write-and-erase performance.

NAND is an ideal solution for high-capacity data storage. Its architecture competes by offering extremely high cell densities that translate to high storage capacity, combined with fast write and erase rates. The cell size of NAND flash is almost half the size of a NOR cell. The difficulty using NAND revolves around the need for flash management and special requirements for the system Static Random Access Memory (SRAM) interface. The NAND technology has enough address pins to map its entire media, allowing for easy access to every byte contained in it. NAND memories are accessed serially via a complicated input/output (I/O) interface, which may vary from one device or vendor to another. In nearly all cases, the same eight pins are used to convey control, address, and data information.

Using a NOR-based flash is a straightforward process. It is connected like other memory devices, and, as noted above, code can be run directly from it. NAND, however, is complicated with its requirement for a relatively sophisticated I/O interface. Accessing rules for NAND interfaces may differ depending on the NAND vendor. A driver must be written and used for performing any operation on a NAND memory. Writing information to a NAND device is a particularly tricky issue, since the designer must not write to a bad block—meaning that virtual mapping must be implemented on NAND devices at all times.

1.2 Charge Storage: Floating Gate versus Discrete Storage

There are a number of emerging technologies based on new materials and/or storage concepts. Both Silicon-Oxide-Nitride-Oxide-Silicon (SONOS) [5-6] and Nitrided Read Only Memory (NROM) technologies [7-8] use a nitride layer, instead of a floating gate, between the control gate and substrate to store charges.

A SONOS memory cell is formed from a standard polysilicon NMOS transistor with a layer of silicon nitride inserted inside the gate oxide. The nitride layer is non-conductive but contains a large number of charge-trapping sites able to hold an electrostatic charge. The layer is electrically isolated from the surrounding transistor, although charges stored on the nitride directly affect the conductivity of the underlying transistor channel.

The NROM cell is comprised of a nitride layer, surrounded by two insulating oxide layers. The nitride layer serves as a trapping dielectric for two separate localized charge packets at each end of the cell to store two bits. Each charge can be maintained in one of two states, either "programmed" or "erased," represented by the presence or absence of a pocket of trapped electrons. This enables the storage of two bits of information without the complexities associated with multi-level-cell technology.

In theory, SONOS and NROM offer higher quality charge storage due to the smooth homogeneity of the nitride film compared with a polycrystalline film. They are also less susceptible to oxide defects. If one leakage path is generated in the bottom oxide, only a few traps will lose charge and, therefore, stress-induced leakage current (SILC) may not be an issue. Compared to floating gate technology, SONOS and NROM are free of drain turn-on and floating gate interference and immune to SILC when the bottom oxide layer is thinner than the floating gate technology. This provides lower program/erase voltages and faster programming. However, every technology has its weaknesses. Both SONOS and NROM have erase saturation (i.e., hard-to-erase and/or over-erase) and the trap-related reliability issues for these two technologies are still not fully understood.

Section 2. Flash Memory Reliability

Flash memory reliability is approaching its ultimate physical limitations. The reliability assessment and prediction for non-volatile memories in such a transition scenario forces one to employ complicated reliability physics in which the support of experimental observation, physical interpretation, and numerical modeling play an unprecedented role. Major reliability concerns for flash memory technologies are endurance, data retention, bit flipping, and bad-block handling [2-9].

2.1 Bit Flipping

All current flash architectures suffer from "bit flipping," when a bit either gets reversed or is reported reversed. Problems associated with bit flipping are more common with NAND devices than with NOR devices; therefore, an error detection code / error correction code (EDC/ECC) algorithm is strongly recommended for NAND devices. This problem is not as critical when using NAND to store multimedia information. However, when used as a local storage device storing the operating system, configuration files, or other sensitive information, an EDC/ECC system must be implemented to ensure reliability.

2.2 Bad-Block Handling

NAND devices are shipped with bad blocks randomly scattered within them. An early attempt to ship NAND devices free of bad blocks was found not to be economically feasible due to the very high price tag caused by low-production yield rates. NAND devices require an initial scanning of the media for bad blocks, which are mapped as unusable. Failing to perform this process in a reliable way results in high failure rates of the finished devices.

2.3 Endurance

Endurance refers to the maximum possible allowed number of erase/write cycles for a memory device. Typically, the maximum allowed number of erase cycles for a NAND device is one million cycles, compared with 100,000 cycles for a NOR device. In addition to the 10-to-1 block-erase-cycle advantage of NAND memory devices, the typical NAND block size itself is approximately eight times smaller than that of a NOR device, which means each NAND memory block will be erased fewer times over a given period of time.

2.4 Retention

Data retention refers to the memory's ability to retain data. A data retention failure is when there is at least 1 bit of data that cannot be read or is read incorrectly. Data retention is also a function of the number of erase/write cycles.

2.5 Data-Loss Effects

Figure 1 shows the different data-loss effects for endurance and data retention. Main V_t loss is believed to result from oxide tunneling and interface/bulk de-trapping. The intrinsic tunneling phenomenon yields a larger spread of the endurance cycle and retention time distributions. Traps generated at the interface and in the bulk tunnel oxide lead to sub-threshold slope degradation affecting V_t. Tail V_t loss is due to SILC. De-trapping and SILC have different temperature dependencies and activation energies, and therefore, the screening and evaluation of advanced flash memory technologies are much more complicated.

<u>Please note</u> that the above data-loss mechanisms apply for all flash memory cells, whether for intrinsic or extrinsic (weak) cells. The tail distribution due to SILC is of an intrinsic nature for advanced flash memory technologies, which is different from the tail distribution induced by extrinsic or "weak" memory cells in EEPROMs (MNOS technology) as described in the NEPP 2005 final report titled, "Body of Knowledge Guideline Document on Commercial MNOS EEPROM in Space Applications" [10].

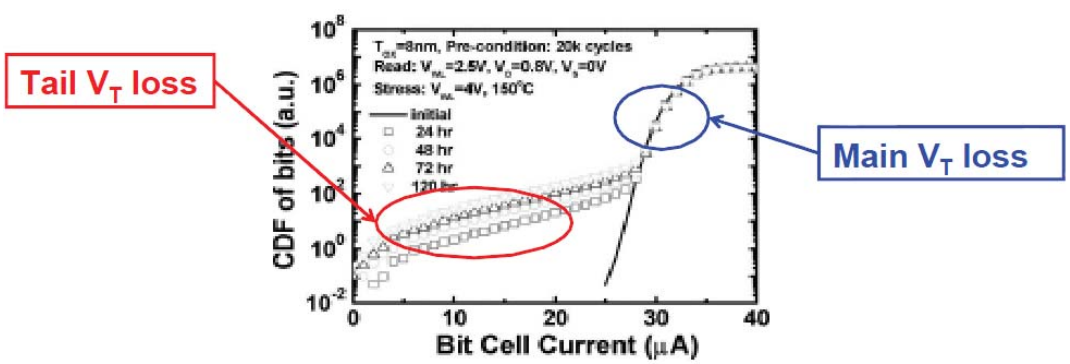

Figure 1. Data-loss effects for flash memory devices.

Section 3. Reliability Testing on Samsung 8-Gbit NAND Multi-Level Cell (MLC) Flash Memory

Samsung 8-Gbit NAND multi-level cell (MLC) flash memory chips were selected for reliability and radiation evaluation in 2008. This flash memory is fabricated using 60-nm technology and its performance includes an 800-μs program time and a 1.5-ms erase time.

The Samsung datasheet indicates that the flash memory properties include 5K program/erase cycles with 4-bit/512-byte ECC, and a 10-year data retention life. Reliability testing, including endurance and data retention tests, was performed on the Samsung 8-Gbit NAND flash memory chips. The test results are summarized below.

3.1 Endurance

Endurance testing was performed without ECC to investigate the raw endurance characteristics of the flash memory chips. Based on the total time to complete 10,000 program/erase cycles, it was decided to conduct the endurance testing on eight blocks; each block includes 128 pages and 1 page has (256K+8K) bytes. The endurance testing was performed at room temperature to investigate the combined effects of de-trapping and SILC mechanisms on endurance characteristics.

Two parts were tested and bit failures were observed for both parts at as low as 1000 program/erase cycles, as shown in Figure 2. In order to investigate the nature of the dominant bit failure mechanisms (i.e., either Time Dependent Dielectric Breakdown [TDDB], charge trapping/de-trapping, and/or SILC), the same parts were tested for an additional 10,000 program/erase cycles at the same bias and temperature conditions for 19 days after completion of the first endurance testing. The parts were kept at room temperature during these 19 days.

Figure 3 shows that bit failures occurred earlier during the second endurance test, done after 19 days, compared to the first endurance test. Table 2 gives the locations/addresses of the failed bits, indicating that the bit failures were consistent during the first and second endurance testing, but did not seem to be permanent between the first and second endurance tests.

Since TDDB typically generates a leakage path through the tunnel oxide and creates a short path between the control gate and substrate, the bit failures due to TDDB should typically be a stuck “0” and should result in initial bit failures at the beginning of the second endurance testing. The fact that no initial bit failures were present at the second endurance testing seems to indicate that the dominant failure mode of the bit failures during the endurance testing was not TDDB, but rather trapping/de-trapping or SILC of the tunneling oxide (or a combination of both).

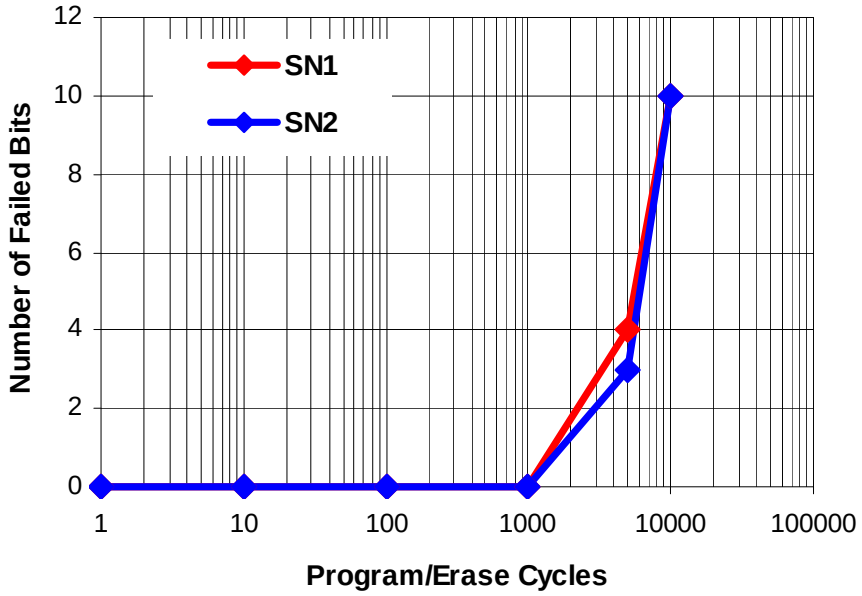

Figure 2. Initial endurance testing results on Samsung 8-Gbit NAND.

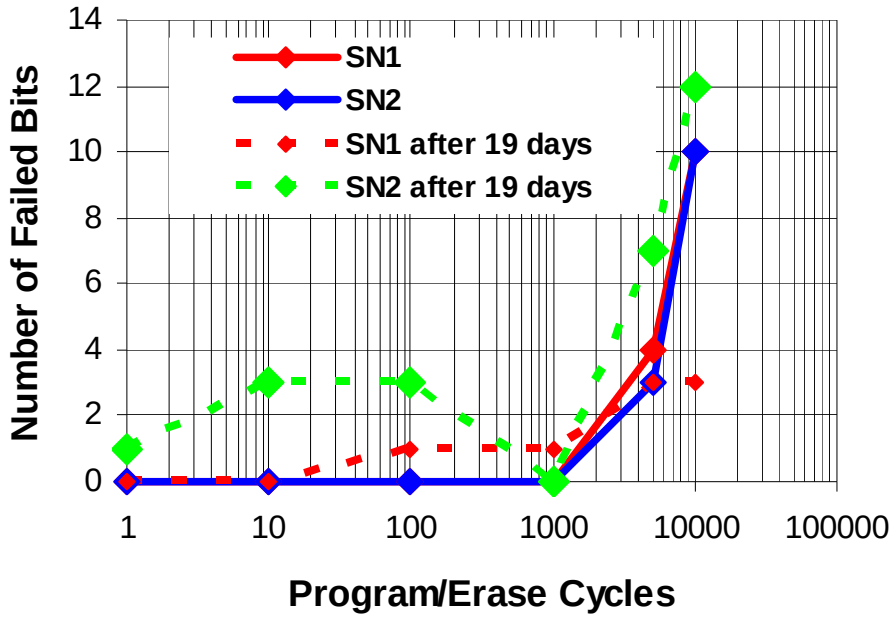

Figure 3. Endurance testing results on Samsung 8-Gbit NAND—initial and after 19 days.

Table 2. Endurance Failure Locations for Samsung 8-Gbit NAND.

SN1	Endurance performed on fresh device					
	Number of Cycles					
Block	1	10	100	1000	5000	10000
1	0	0	0	0	5/1577/8	5/1577/8
2	0	0	0	0	4/693/40, 4/1002/20	4/693/40, 4/1002/20
3	0	0	0	0	0	0
4	0	0	0	0	0	0
5	0	0	0	0	0	4/832/40
6	0	0	0	0	0	0
7	0	0	0	0	5/589/4	5/589/4, 5/389/80
8	0	0	0	0	0	1/299/80, 5/299/80, 5/2096/1

SN1	Endurance performed 19 days after first 10K write/erase cycles					
	Number of Cycles					
Block	1	10	100	1000	5000	10000
1	0	0	0	0	0	0
2	0	0	0	0	0	5/1061/10
3	0	0	0	0	0	0
4	0	0	0	0	5/692/10	0
5	0	0	5/589/4	5/589/4	5/589/4	0
6	0	0	0	0	0	0
7	0	0	0	0	0	5/1313/4
8	0	0	0	0	5/299/80	5/299/80

0	zero bits failures
5/1577/8	failed bit's address or location

3.2 Retention

Data retention testing was performed on two fresh Samsung 8-Gbit NAND flash memory devices. Continuous read cycles were conducted at 125°C during the data retention testing. Characterization was done at room temperature at retention time intervals at zero, 100 hours, and 500 hours. One bit failure at the same location was first observed on one flash device at the 100-hour interval, compared to the 10-year data retention specifications given in the datasheet.

3.3 Summary

Early bit failures were observed during both endurance and data retention testing, but it should be noted that no ECC was applied during the testing. However, both endurance and data retention life specifications are defined with ECC in the datasheet. ECC is a MUST for advanced flash memory technologies for any space applications.

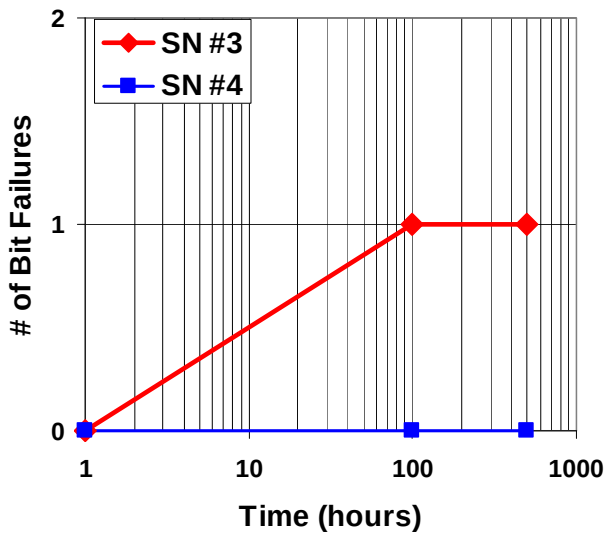

Figure 4. Data retention testing results on Samsung 8-Gbit NAND.

Section 4. Preliminary Thoughts on Risk Mitigation

Reliability and radiation evaluation of advanced flash memory technologies has just begun and has not yet been completed for this task. Below is a list of preliminary thoughts on risk mitigation for flash memory technologies potentially used for space applications.

4.1 Reliability

Flash memory technologies can exhibit a number of possible reliability mechanisms related to program/erase cycling and retention. Managing these mechanisms involves technology optimization as well as media management. Qualifying these devices requires a range of stresses to target the full set of failure mechanisms. “Standard” qualification and screening flows can be defined, but knowledge-based methods will always be needed for specific applications. This requires a thorough understanding of the failure mechanisms and detailed analysis of the application specifics [2-9].

4.2 Redundancy

Risk mitigation for flash memory applications needs to be implemented at both part and system/design levels. Therefore, code and/or image redundancy should be required to improve flash memory reliability in applications.

All flash memories should utilize ECC, which requires adding extra parity bits to a page sufficient to reconstruct the correct data when a certain number of bits are in error. In this

manner, ECC can dramatically reduce the bit error rate. However, any ECC scheme will fail if the raw bit error rate is too high, as illustrated in Figure 5. For this reason, ECC is most effective for "thin tail" mechanisms like SILC and least effective for intrinsic shifts such as that caused by de-trapping. In general, the amount of ECC needed must be carefully chosen based upon reliability characterization data [11-13].

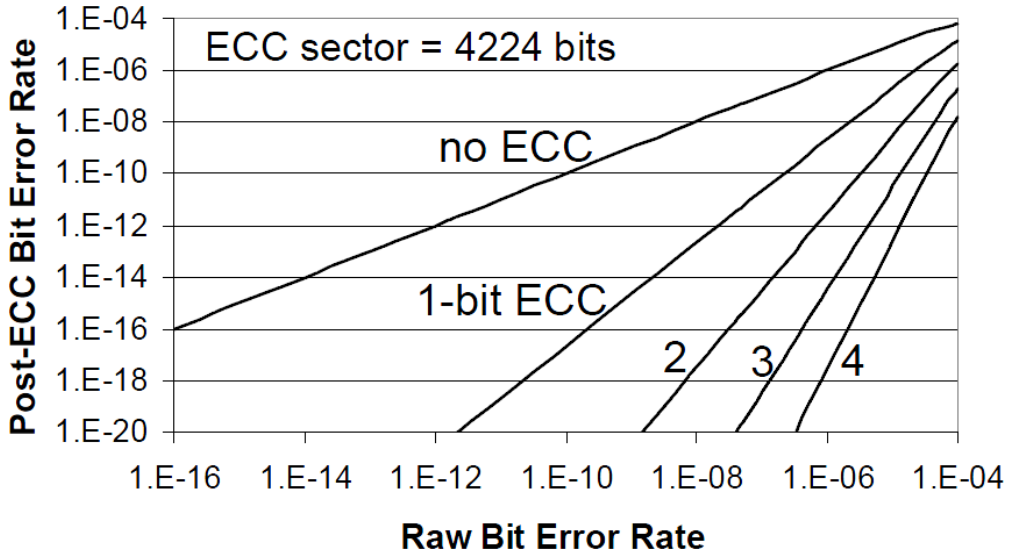

Figure 5. Effect of ECC.

4.3 Media Management

NAND devices usually have the required ECC specified on their datasheets, but the actual ECC is performed by the system. In addition to ECC and redundancy applied at the part and system level to improve overall flash memory reliability, media management approaches should be implemented to manage mapping from logical to physical addresses, to detect and map out bad blocks, and to perform wear-leveling for flash memory life extension [12-13]. For example, program/erase trap-up can be managed by intelligent pulse/verify/repeat algorithms, and erratic erase can be managed by post-erase-repair algorithms. Many flash failure mechanisms worsen with cycling. As a result, many systems implement wear-leveling, in which the system moves often-changed data from one block to another to prevent any one block from getting far more cycles than others.

References

[1] International Technology Roadmap for Semiconductors, 2008.

[2] R. Bez, et al., "Introduction to Flash Memory," *Proceedings of the IEEE*, 2003.

[3] M. Gill, "Flash Memories: A Review," *IEEE International Nonvolatile Memory Technology Conference*, 1996.

[4] A. Modelli, et al., "Flash Memory Reliability," *IEEE Integrated International Reliability Workshop*, 2005.

[5] K.-H. Wu, et al., "SONOS Device with Tapered Bandgap Nitride Layer," *IEEE Transactions on Electron Devices*, Vol. 52, 2005.

[6] T.-S. Chen, et al., "Performance Improvement of SONOS Memory by Bandgap Engineering of Charge-Trapping Layer," *IEEE Electron Device Letters*, Vol. 25, 2004.

[7] B. Eitan, et al., "NROM: AS Novel Localized Trapping, 2-bit Nonvolatile Memory Cell," *IEEE Electron Device Letters*, Vol. 21, 2000.

[8] A. Shappir, et al., "The Two-bit NROM Reliability," *IEEE Transactions on Device and Materials Reliability*, Vol. 4, 2004.

[9] C.C. Yeh, et al., "Reliability and Device Scaling Challenges of Trapping Charge Flash Memories," *Proceedings of the 11th International Symposium on the Physical and Failure Analysis of Integrated Circuits*, 2004.

[10] Y. Chen, *Body of Knowledge Guideline Document on Commercial MNOS EEPROM in Space Applications*, 2006.

[11] N. Mielke, "NVM Reliability," *International Reliability and Physics Symposium*, 2007.

[12] D. Ielmini, "Physical Mechanisms and Modeling of Flash and Post-Flash Reliability," *International Reliability and Physics Symposium*, 2008.

[13] Y.-J. Choi, "Flash Memory Reliability and its Management from System Perspective," *International Reliability and Physics Symposium*, 2008.