National Aeronautics and Space Administration

NAND Flash Screening and Qualification Guideline for Space Application

Jason Heidecker
Jet Propulsion Laboratory
Pasadena, California

Jet Propulsion Laboratory
California Institute of Technology
Pasadena, California

JPL Publication 12-1 2/12

National Aeronautics and Space Administration

NAND Flash Screening and Qualification Guideline for Space Application

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

Jason Heidecker
Jet Propulsion Laboratory
Pasadena, California

NASA WBS: 724297.40.49
JPL Project Number: 104593
Task Number: 40.49.01.04

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.

Table of Contents

1 – Introduction **1**
2 – Reliability Issues **1**
2-1 NAND Flash Overview **1**
2-2 Flash Reliability **2**
Bad Blocks 3
Data Retention and Endurance 3
Disturb Errors 4
2-3 Failure Mechanisms **4**
2-4 Scaled CMOS Failure Mechanisms **5**
Electromigration (EM) 7
Time Dependent Dielectric Breakdown (TDDB) 7
Hot Carrier Injection (HCI) 7
Negative Bias Temperature Instability (NBTI) 7
2-5 Plastic Packaging Failure Mechanisms **7**
2-6 Floating Gate Failure Mechanisms **8**
Stress Induced Leakage Current (SILC) 8
Dielectric Trapping and Detrapping 8
Surface Inversion (Mobile Ions) 8
2-7 NAND Architecture Failure Mechanisms **8**
Disturb Errors 8
Read Disturb 9
Program Disturb 9
2-8 Failure Modes **9**
3 - Screening and Qualification **10**
3-1 EEE Device Reliability Overview **10**
3-2 Screening and Qualification **10**
3-3 Test Flows **11**
Failure Criteria 13
3-4 Reliability Characterization **13**
Endurance Cycling 14
Data Retention Bake 15
Read Disturb 15
UBER Calculation 15
3-5 Radiation Characterization **17**
4 – System Level Considerations and Other Recommendations **18**
4-1 Error Detection and Correction (EDAC) **18**
4-2 Wear Leveling **18**
4-3 Other Considerations **18**
Part Selection, Pure Tin Leads 18
Disturb Error Mitigation 18
Moisture Sensitivity, Proper PEM Storage 19
Electrostatic Discharge (ESD) Sensitivity 19
5 – Conclusion **19**
6 – References **19**

NAND Flash Memory Qualification Guideline

1 – Introduction

All space missions have a need for nonvolatile memory (NVM) that maintains data integrity when unpowered. Types of NVM include PROM, EEPROM, NOR Flash, and NAND Flash. PROMs, NOR Flash, and EEPROMs are good choices for storing smaller file size data such as boot code or FPGA configurations. These products have excellent data retention characteristics and can reliably store data for years, unpowered, without any data corruption. They can also be read many times without disturbing the data. Another application of NVM is storage of science and engineering data, which requires large of amounts of memory. The highest density memories available today are SDRAM and NAND Flash. However, the power required to operate and store data in NAND Flash is far less than SDRAM. Wherever high density and low power is required, NAND Flash is very attractive.

Unfortunately, high density NAND Flash memory is a completely commercial product, from its commercial CMOS processes to its plastic packaging. It is not offered as a DLA (Defense Logistics Agency) qualified product or fabricated on any radiation-hardened process lines. Therefore, these products face a greater barrier to entering the space market than other memories like SRAM, which is fully qualified by the DLA and is offered with hermetic packaging and rad-hard CMOS. Some NAND products are available with radiation effects mitigation built in, but the selection is limited [1]. When these products are not suitable for a particular application, the space project must look to other commercial NAND products, screening and qualifying them for reliability and radiation effects.

To date, many missions have shown an interest in utilizing high density NAND Flash for space applications [2]. The purpose of this guideline is to provide space missions with background information and a qualification methodology for commercial NAND Flash products. First, the primary failure modes and mechanisms are described, along with the tests required to mitigate against them. Most testing of this type is covered by the manufacturer's technology or product qualification flow, making it the space project's responsibility to independently verify the manufacturer's claims, perform additional reliability and radiation qualification testing, and finally screen flight units. Finally, this guideline also provides information regarding selection, handling, and system-level implementation issues.

2 – Reliability Issues

2-1 NAND Flash Overview

NAND Flash is a nonvolatile memory solution that offers the highest densities of any memory on the market. NAND Flash is available up to 512 Gb built on 20-nm commercial CMOS processes. The most popular package types are thin-small-outline-package (TSOP), ball-grid-array (BGA), and land-grid-array (LGA). All are plastic-encapsulated-microcircuits (PEMs).

The two types of Flash are NAND and NOR, which are distinguishable by the way their memory cells are connected. In NOR, cells are connected in parallel to the bit lines, resembling a NOR gate (Fig. 1); this structure allows random access to each cell. In NAND, cells are connected serially, resembling a NAND gate (Fig. 1). In this configuration, connections between each cell and the bit line are removed, reducing cell size but eliminating random access. This architecture is what gives NAND its density advantage over NOR: NAND has cell sizes of $4F^2$ and NOR's are $10F^2$.

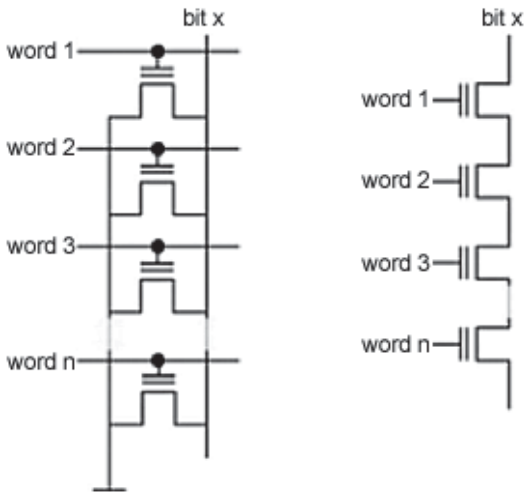

Fig 1. Comparison of NOR (left) and NAND (right) Flash cells.

Flash users interact with these memory arrays through a few basic commands: Read ID, Erase Block, Program Page, Read Page, and Check Status. The Read ID simply returns a few bytes of data identifying basic information about the device, including architecture, manufacturer, and part number. Erase, Program, and Read are self-explanatory. Check Status is a simple command issued after other commands, running it causes a byte of data to be returned that identifies whether or not the previous command was successful. This is an important operation, and should be performed after every Program and Erase command; the "Bad Blocks" section below will discuss this further. Any NAND Flash datasheet can describe these commands in more detail. There are complicated ways of performing these operations such as interleaved die operations, but these five operations are the basics.

Flash memories are partitioned into blocks, which are further partitioned into pages. Pages are usually a few dozen kilobytes in size. Flash memory cells are essentially a standard MOS transistor with a floating gate between the transistor channel and control gate. By moving charge on and off this floating gate, the threshold voltage of the memory cell can be changed, thus changing the state/data of the memory cell. Erase operations are done at the block level while program and read operations are done page by page. Both erase and program operations are accomplished by moving electrons on and off a floating gate in the cell by a process known as Fowler-Nordheim tunneling (FNT), which involves high voltages (12-20 V) and high electric fields. FNT is destructive to the cell, in effect consuming some of its life with each Erase or Program operation. FNT, along with NAND's unique serial cell arrangement and the scaled-CMOS it is built on, form the basis of the primary reliability concerns with NAND Flash.

2-2 Flash Reliability

Discussions regarding Flash's reliability tend to center around four topics:

1. Bad Blocks
2. Data Retention
3. Endurance
4. Disturb

Failure in Flash memories has a slightly different meaning than with other types of memory. This is because the dominating failure mechanisms are unique to Flash memories, and are unrelated to the underlying CMOS technology. Reliability of NAND Flash is dominated by two types of failures: 1) those related to the floating gate properly holding its charge (retention and endurance), and 2) disturb phenomena related to NAND's unique cell arrangement and architecture.

The floating gate failures are related to the destructive FNT used to program and erase memory cells, which will consume the life of the memory array long before any packaging or CMOS related failure mechanism starts to degrade performance. The underlying CMOS generally has a 10-year lifetime at 55°C, whereas a few thousand program/erase cycles of a device can render it unusable once the cells' gate oxides fail to reliably hold a charge.

The data corruption caused by disturb phenomena are independent of any intrinsic material limitations or wear-out of the device. Disturb issues are the result of NAND's cell arrangement and architecture, which is more susceptible to disturb errors than other memories.

Bad Blocks

NAND Flash devices are expected to be defective as delivered right from the factory. In order to increase yield in a highly competitive commercial market, NAND Flash manufacturers take advantage of the physically isolated block structure by marking blocks with defects as "bad blocks" and declaring the rest of the device fit to use. Bad blocks have specific data patterns programmed into the spare byte area of their pages. This way users can identify the bad blocks and avoid them. Also, more blocks can "go bad" over the life of the part. This manifests itself to the user as a failed block erase or failed page program. Any time the Check Status command following these operations returns a failure, these blocks should be marked as bad and mapped out of the usable address space. Also, there is a third way the manufacturer defines a bad block: whenever a block fails the specified error correction code criteria. For example, "12 bit ECC per 539 byte sector up to 5,000 program/erase cycles" is the specification for a 32 Gb device from Micron. So, any time the user sees 12 bit errors in a 539 byte sector read within the first 5,000 program/erase cycles, that block should be marked as bad.

Manufacturers specify in the datasheet the maximum number of bad blocks the part can have, which is 2.5% for the 32 Gb device mentioned above. Since this device has 8,196 blocks, up to 204 could go bad by the time 5,000 program/erase cycles have been performed (this assumes all 8,196 blocks are being cycled).

Data Retention and Endurance

Although NAND Flash is a nonvolatile memory, the memory cell technology is not perfect and the charge on any floating gate will eventually leak off. This process is accelerated with decreasing temperature and is known as a "data retention" error. These are 'soft' errors and once reprogrammed, the cell will again hold their charge as expected. Again, the ECC specification comes into play in order to determine data retention failure: For the 32 Gb device previously discussed, the data retention spec means the user should never see 13 or more bit errors in any 539 bytes read after 5,000 cycles and 10 years at 55°C.

When a cell is erased and programmed via FNT too many times, an endurance failure will occur. This means the high electric fields involved with FNT have degraded the oxide to the point that it will no longer properly program or erase the cell. This is known as an endurance failure, a 'hard' failure in which the cell becomes stuck at 1 or 0.

The physical mechanisms behind retention and endurance failures are known as time-dependent-dielectric-breakdown (TDDB), stress-induced-leakage-current (SILC), dielectric trapping and detrapping, and surface inversion (mobile ions). All of these mechanisms and more are discussed in Sections 2-3 thru 2-7.

Disturb Errors

All memory technologies, not just NAND Flash, suffer from disturb phenomena. However, some are more sensitive to certain types of disturb than others. A disturb error means that the data in one cell is corrupted due to operations performed on another cell. This happens when the voltages and electric fields involved in reading, erasing, and programming inadvertently change the threshold voltages (and therefore data) in unselected cells.

Erase disturb is not a problem in Flash memories. Erase operations occur at the block level, and erase voltages are localized to the targeted block only.

Program and read disturb errors do occur in NAND Flash, and some types of NAND Flash are more susceptible to read and program disturb errors than others.

To better understand reliability related to this phenomenon, it is important to point out that NAND Flash comes in two major varieties: Single-level cell (SLC) and multi-level cell (MLC). As the names suggest, SLC devices store one bit per cell and MLC devices store 2 or more bits per cell. In SLC devices, memory storage is accomplished by changing the threshold voltage of the cell transistor by adding charge to a gate above the transistor channel. The two states (gate charged or uncharged) change the on/off characteristic of the cell and the sensing circuitry then detects the states as a 1 or 0. MLC devices follow the same concept, except four voltage levels (for 2-bit cells) or eight voltage levels (for 3-bit cells) are recognized by the sensing circuitry (Fig. 2).

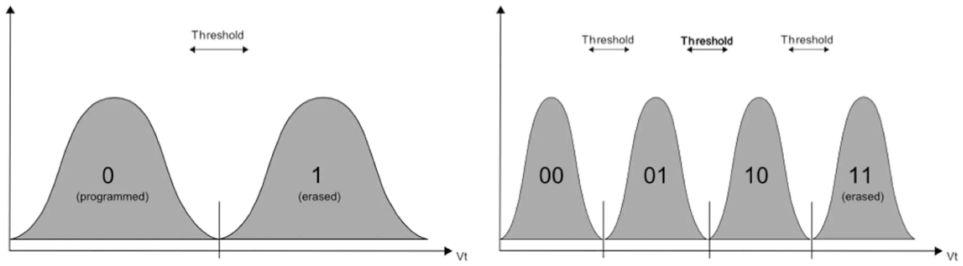

Fig. 2. Voltage margins for SLC (left) and MLC (right) NAND Flash.

Because the voltage margins in MLC devices are smaller than SLC, they are more susceptible to disturb phenomena. In MLC devices, it takes a smaller amount of charge, moved on or off the floating gate, to corrupt data. Disturb errors of any kind are very rare in modern SLC devices and maybe never be experienced in the field at all [3]. However, in MLC devices, disturb errors can happen on the very first operation of a virgin device, and the bit error rates due to this phenomena get worse with program/erase cycling.

2-3 Failure Mechanisms

NAND Flash devices, like any EEE component, suffer from a broad range of extrinsic and intrinsic failure mechanisms. The distinction between intrinsic and extrinsic failures is important, because it points to possible causes of failure, and therefore indicates the direction of preventive or corrective actions. The failure mechanisms discussed in this section are what lead to the major reliability issues discussed above, such as endurance, retention, disturb, bit errors, and bad blocks.

Extrinsic failure mechanisms are those caused by workmanship, environment, and variability in manufacturing processes. These include fabrication defects, mishandling, electrical overstress, and radiation upset. Extrinsic failure mechanisms are typically combated by visual inspection (checking for chip outs, corrosion, bent leads), screening (X-ray, burn-in, temperature cycling), and special care during handling (ESD precautions, dry boxes).

Intrinsic failure mechanisms are unrelated to workmanship and are inherent to the technology or materials used to create the product. For example, no matter how perfectly a gate oxide is grown, application of voltages and electrical fields across the oxide will eventually lead to dielectric breakdown and failure. Intrinsic failure mechanisms are often categorized by the processes that drive them: thermal, electrical, or chemical.

Fig. 3 gives an overview of the four major technology areas of NAND Flash. When considering these devices for space application, a comprehensive qualification approach must address the failure mechanisms and reliability issues of all four areas: highly scaled CMOS, the floating gate memory storage element, plastic packaging, and device architecture.

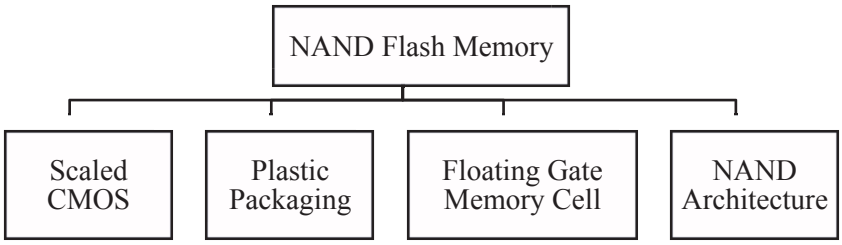

Fig. 3. Areas of focus for qualifying NAND Flash.

Although both CMOS- and packaging-related failure mechanisms are carefully considered by the manufacturer (who must design in reliability and have good yield), it is the floating gate and architectural-related failure mechanisms that dominate reliability in the field and must be the priority of space application qualification efforts.

Table 1 is a summary of the major failure mechanisms in these four areas along with the type of test (qualification, screening, or other) used to mitigate against the failure mechanism or verify that the risk of failure is low.

2-4 Scaled CMOS Failure Mechanisms

CMOS feature sizes continue to shrink with each generation in order to produce the highest performing, most capable integrated circuit that technology can offer. Shrinking the size of the transistors and other CMOS elements has effects on circuit design, performance, power consumption, and reliability.

The four dominant failure mechanisms related to scaled CMOS are electromigration (EM), TDDB, hot-carrier-injection (HCI), and negative-bias-temperature-instability (NBTI) [4]. All of these failure mechanisms can be tested through standard life testing. However, because HCI has a negative activation energy, both high temperature and low temperature operating life testing is required. Also, all CMOS processes are susceptible to defects (for which temperature and voltage accelerated burn-in is a good screen).

Table 1. NAND Flash Failure Mechanisms

Failure Mechanism	Technology Area	Mitigation/Verification		
		Qualification (Sample)	Screening (100%)	Other
Electromigration	CMOS	HTOL		
Time Dependent Dielectric Breakdown (TDDB)	CMOS	HTOL		
Hot Carrier Injection (HCI)	CMOS	LTOL[1]		
Negative Bias Temperature Instability (NBTI)	CMOS	HTOL		
Material Defects	CMOS		Burn-In	
Corrosion	Packaging			Proper material selection
Mechanical Failure (Thermal Cycling)	Packaging	Extended temperature cycling		Proper material selection, Tg measurement, operate below Tg
Voiding, Delamination, Popcorning	Packaging		X-ray/CSAM[2]	Pre-conditioning (bake-out), proper PEM packaging and storage
Tin Whisker Growth	Packaging			Avoid pure tin
Stress Induced Leakage Current (SILC)	Floating Gate	Endurance cycling, Data retention bake		
Dielectric Trapping and Detrapping	Floating Gate	Endurance cycling, Data retention bake		
Surface Inversion, Mobile Ions	Floating Gate	Endurance cycling, Data retention bake		
Read Disturb	NAND Architecture			Read disturb characterization, limits reads between programs
Program Disturb	NAND Architecture			Program disturb characterization, use sequential page programs, avoid partial page programs

[1] Typically performed during technology or product qualification and not required during upscreening.
[2] Due to advances in modern packaging processes, the current NASA JPL PEMS Guideline does not require X-ray or CSAM screening [5].

Electromigration (EM)

An EM failure is the result of high current density creating a transport of bulk material down the metal line. The metal line begins to thin out (increased resistance) and eventually will be turned into an open circuit.

As device features get smaller with scaling CMOS, so do the dimensions of metal lines. Unfortunately, power consumption is also increasing with newer technologies. Smaller metal lines and higher currents mean increased current density through the metal layers and increased risk of EM failure.

Time Dependent Dielectric Breakdown (TDDB)

TDDB refers to the relatively slow degradation of insulating oxides repeatedly exposed to electric fields over time. At nominal logic voltage levels, this process is so slow it never becomes a practical concern. However, in Flash memory cells where high electric fields are used in FNT during program and erase operations, TDDB becomes a real concern.

Silicon dioxide, SiO_2, is used as the primary insulating material in modern integrated circuits. In a floating gate memory cell, it acts at the barrier between cells, gates, contacts, etc. When electric fields are applied to these oxides, electric charges with sufficient energy can tunnel into the oxide and become trapped in the bulk of the material or at the interface.

This trapped charge also increases the local electric field, creating a positive feedback that leads to accelerated charge collection until the tunneling current is large enough to rupture the gate oxide.

Hot Carrier Injection (HCI)

HCI describes the phenomenon by which carriers gain sufficient energy to be injected into the gate oxide. This occurs as carriers move along the channel in MOSFET and experience impact ionization near the drain end of the device. The damage can occur at the interface, within the oxide and/or within the sidewall spacer. Interface-state generation and charge trapping induced by this mechanism result in transistor parameter degradation, typically switching frequency degradation, rather than a 'hard' functional failure.

HCI is one of the few IC failure mechanisms that has a negative activation energy, and therefore gets worse with decreasing temperature, rather than increasing temperature.

Negative Bias Temperature Instability (NBTI)

NBTI is a wear-out mechanism experienced by PMOSFETs with the channel in inversion. It is believed that NBTI is controlled by an electrochemical reaction, where holes in the PMOSFET inverted channel interact with Si compounds at the Si/SiO_2 interface to produce donor type interface states and possibly positive fixed charge. NBTI damage is generated by cold holes (thermalized) in the inverted channel. The NBTI damage may lead to substantial PMOSFET parameter changes, in particular to an increase of the absolute value of the threshold voltage (transistor is harder to turn on), as well as mobility degradation with consequent reduction in drive current.

2-5 Plastic Packaging Failure Mechanisms

All NAND Flash products are packaged as plastic encapsulated microcircuits (PEMs). The primary concerns with plastic packaging are heat conduction, outgassing, moisture absorption, voiding, delamination, thermal coefficient mismatches, pure tin leads and tin whiskers, and exceeding the glass transition temperature (T_g) of the epoxy [5].

To mitigate against these failure mechanisms, it is necessary to: 1) stay below the glass transition temperature during storage and operation; 2) use CSAM to check for voiding and delamination; 3) bake-out the device at high temperature prior to assembly (known as pre-

conditioning); and 4) select devices that only use epoxy known to be free of contaminants, as well as capable of surviving extended temperature cycling without damaging the lead frame or internal wirebonding. Finally, devices with pure tin leads must be re-tinned or solder dipped in a way to completely cover all pure tin surfaces. This will ensure that tin whiskers cannot grow and possibly create shorts across leads or other board signals.

2-6 Floating Gate Failure Mechanisms

Stress Induced Leakage Current (SILC)

A result of TDDB in the floating gate oxide of a Flash cell is increased low-field leakage current, known as stress induced leakage current (SILC). SILC is the driver of data retention errors and eventually endurance errors. As leakage current increases, the cell's ability to retain the charge needed to significantly shift its threshold voltage is diminished. These are known as data retention errors. Although, keep in mind, even virgin NAND cells programmed only one time have a limited retention characteristic. Even the most perfectly fabricated floating gate and oxide will slowly leak the charge away. SILC exacerbates the effect. Eventually, after the cell has been programmed and erased thousands or millions of times (depending on technology), the gate oxide no longer acts as an insulating material and hard failure of the cell will occur.

Standard endurance cycling and data retention testing can verify that this effect is not a risk to the device's performance above and beyond mission requirements.

Dielectric Trapping and Detrapping

Flash cells store data in the form of charge on the floating gate. The Program command is used to inject electrons and Erase is used to remove electrons. Repeated Program and Erase operations leads to trapped charge on the gate oxide, which affects the threshold voltage of the cell. When the charge detraps over time, the threshold shifts, and this can lead to data loss.

Performing a data retention bake test on devices that have been cycled, say 0.5 or 1.0 times greater than the device endurance specification, can verify that the data retention characteristic will be acceptable at end-of-life.

Surface Inversion (Mobile Ions)

Mobile ions, such as alkaline-metal elements like Li, Na, and K, can contaminate semiconductor processing materials. When these materials build up at the Si/SiO_2 interface, surface inversion can happen and lead to increased cell leakage and device failure.

Again, performing a data retention bake test on devices that have been previously program/erase cycled can verify that the data retention characteristic will be acceptable at end-of-life.

2-7 NAND Architecture Failure Mechanisms

Disturb Errors

All memory technologies, not just NAND Flash, suffer from disturb phenomena. A disturb error means that the data in one cell is corrupted due to operations performed on another cell. This happens when: 1) the voltages involved in erasing, programming, or reading targeted cells changes the threshold voltage in other cells sharing the same voltage lines, or 2) these voltages create electric fields that slightly change the threshold voltage in neighboring cells.

Erase disturb is not a problem in Flash memories. Erase operations occur at the block level, and erase voltages are localized to the targeted block only.

Program and read disturb errors do occur in NAND Flash, and some types of NAND Flash are more susceptible to read and program disturb errors than others.

Read Disturb

Read disturb errors can happen with the very first read of a virgin device, and the failure mechanism gets worse with program/erase cycling. This happens when the page read operation corrupts data within the page being read or within other pages of the same block. All Flash devices intended for space application should be tested for read disturb errors. This means cycling the device, and performing multiple reads of a given data pattern to see how the bit error rate is affected by multiple reads over device lifetime.

Program Disturb

Program disturb is the unintentional programming of cells other than the targeted cells, such as those on the same bitline or wordline. This occurs as the high voltage required for FNT based program operation is not isolated from the unselected bits. Data corruption can occur when cell states are changed by the program operation in the same page or another page within the block.

Testing for program disturb is not very useful for most applications. In practice, as you program pages in a block, along the way bits in unselected pages will be corrupted. This happens much more frequently in MLC, and may never be detected in a modern SLC sample. However, to detect the program disturb, the pages must be read, and determining whether any bit errors are program or read disturb is not possible. (However, the errors from subsequent reads are certainly read disturb errors).

2-8 Failure Modes

There are three types of failure modes that will occur due to the failure mechanisms discussed above. A failure occurs in Flash whenever an Erase, Program, or Read operation fails. Erase and Program failures are rare, but can happen. Users should issue a Check Status command after an Erase or Program operation to verify its success. If these operations ever fail, then the blocks they operated on should be marked as bad and no longer used.

Verifying Read operations is more complicated and either knowing the data pattern ahead of time or using ECC is the only way to know if data integrity is maintained.

Table 2 summarizes these failure modes, detection methods, and countermeasures.

Table 2. NAND Flash Failure Modes

Failure Mode	Detection Method	Countermeasure
Erase Failure	Check status after erase	Block replacement
Program Failure	Check status after program	Block replacement
Bit Errors (within spec)	Verify ECC	ECC correction
Bit Errors (beyond spec)	Verify ECC	Block replacement

3 - Screening and Qualification

The commercial NAND Flash memory market is a highly commoditized field where reliability and quality are required for any manufacturer to remain competitive. Significant effort is put into technology qualification and designing in reliability. However, due to the high cost of 100% exhaustive testing, very little is done in the way of screening beyond a simple room temperature functional go/no-go test. Therefore it is the responsibility of the space project to complete qualification testing for space application as well as fully screen devices to weed out infant mortals.

3-1 EEE Device Reliability Overview

The purpose of reliability testing is to quantify the expected failure rate of a device at various points in its life cycle. Reliability screens and accelerated aging tests rely on the rapidly declining failure rate of infant mortality failure to reduce the early life failure rate. Burn-in, which is conducted at elevated temperature and voltage, is an example of such a reliability screen. Lifetimes of ICs are predictably modeled by the bathtub curve in Fig. 4.

Region I is characterized by high failure rates due to extrinsic failure mechanisms such as manufacturing defects. The purpose of screening is to move a lot of devices through Region I into Region II. Along the way, the weakest devices will fail and the failure rate of the rest of the population will be much lower than the failure rate of the population at time-zero.

Region II is known as the "useful life" of the population and has a lower and constant failure rate. Product and technology qualification testing ensures that the dashed line between Regions II and III is well beyond the guaranteed lifetime of the device. The space project should then repeat HTOL and temperature cycling on the flight lot and verify that this high-reliability regime in Region II extends beyond mission lifetime.

In Region III the failure rate increases as the reliability of the device becomes dominated by its intrinsic failure mechanisms. These are known as "wear out" failures and mark the end of life of the device population.

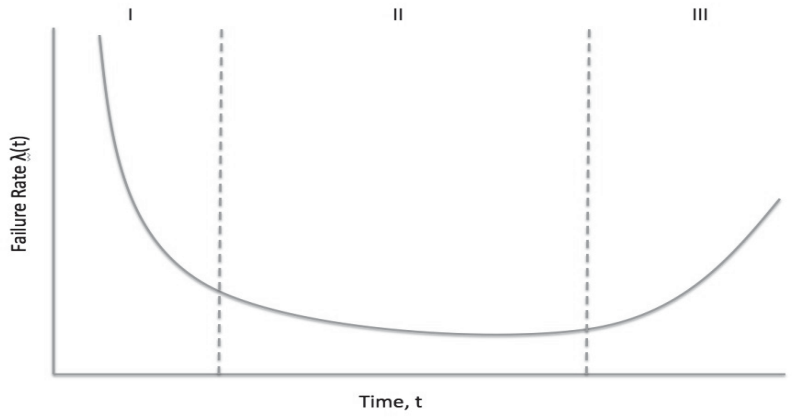

Fig. 4. Bathtub curve modeling lifetime of EEE devices.

3-2 Screening and Qualification

Fig. 5 summarizes the testing involved in screening and qualifying NAND Flash for space. The testing is grouped by reliability type as well as responsible party. The manufacturer is mostly concerned with designing in reliability and proving reliability at the technology level. They want to make sure their devices can operate for many years without failure due to CMOS related failures such as electromigration and TDDB. It is then the responsibility of the space project to verify the

manufacturer's claims with lot-specific qualification followed by radiation characterization, reliability characterization (endurance cycling, read disturb, bit error rate characterization) and, finally, 100% flight unit screening.

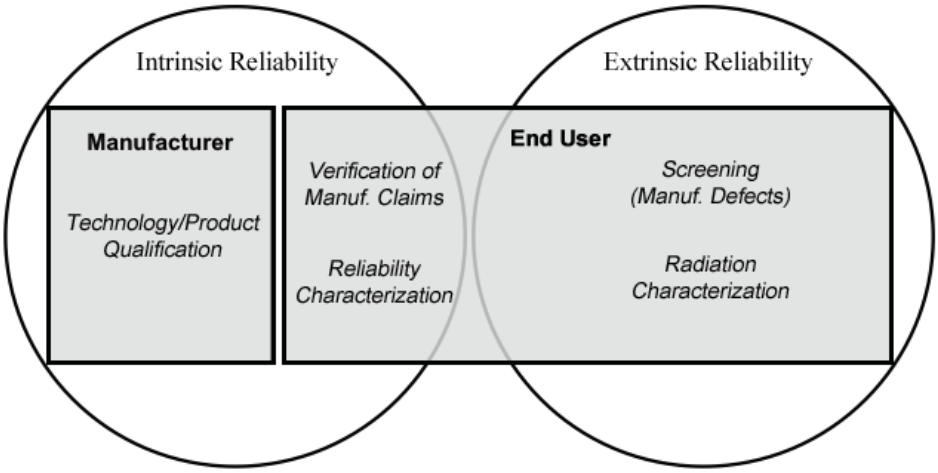

Fig. 5. Diagram depicting the roles of the manufacturer and space project in screening and qualifying an EEE device for space application.

3-3 Test Flows

The goal of screening is to ensure the product enters the field in Region II of its lifetime (see Fig. 4) with a very low likelihood of early-life failure due to manufacturing defects. To save costs, NAND Flash manufacturers do not do 100% screening or burn-in, which are obviously very important to a high reliability application such as a space mission. Therefore the space mission must upscreen all commercial NAND Flash before flight.

The goal of qualification testing is to verify that the transition between Regions II and III of the part's lifetime (see Fig. 4) is well beyond what is required by the space mission. In practice, the manufacturer handles most of the technology qualification, proving their devices can operate at 55°C for 10 years with a very low likelihood of failure due to EM, TDDB, NBTI, etc. It is the responsibility of the space project to verify these claims and perform additional radiation qualification and reliability characterization to demonstrate extrinsic reliability of the device.

Table 3 gives recommended screening and qualification test flows for Level 1 and Level 2 mission profiles as defined by JPL's Part Engineering Technical Standard [6]. It is recommended that Level 1 missions require additional burn-in, lower PDA, and larger sample sizes for life test, DPA, and extended temperature cycling. All qualification samples should have already passed the screening tests as indicated. This table is a comprehensive list of all screening, qualification, and characterization testing required by the project. However, the project should keep in mind this guideline is only a recommendation, and that actual flight qualification and screening test flows should be reviewed by the appropriate parts specialist working for the project.

The flow begins with measuring the glass transition temperature of the epoxy used to encapsulate the microcircuit. This temperature should be avoided during testing and flight. The mismatch between coefficient of thermal expansion of the epoxy and other PEM internal materials above the glass transition temperature will cause significant stress buildup in packaged components during mission applications, possibly leading to early device failures [5].

Table 3. NAND Flash Screening and Qualification Test Flow

Step	Test	Requirements	Sample Size & Notes	
			Level 1 Mission	Level 2 Mission
1	Glass transition temperature, Tg	Thermal Mechanical Analysis (TMA)	3 pc	3 pc
2	Serialization		100%	100%
3	Electricals (AC, DC, Functional)	Test to datasheet. Read & record data. Tri-temp (-40°C, 25°C, 85°C).	100%	100%
4	Dynamic Burn-In	MIL-STD-883, Method 1015, 125°C	100% 240 hrs	100% 160 hrs
5	Electricals (AC, DC, Functional)	Test to datasheet. Read & record data. Tri-temp (-40°C, 25°C, 85°C).	100%	100%
6	Delta Calculations	25°C	100%	100%
7	Percent Defective Allowable (PDA) Calculation		5% PDA[1]	5% PDA
8	Stabilization Bake	125°C, 24 hours	100%	100%
9	Flight Part Storage	In accordance with [5].	100%	100%
10	Split Lot	22 pc (temp cycle), 22/45 pc (life test), 3/5 pc (DPA) 10 pc (application-specific qual)	n/a	
11	Preconditioning	JEDEC 22A-113F	All temp cycle and life test samples	
12	Temp Cycling	MIL-STD-883, Method 1010, Condition B	22 pc, 300 cycles	22 pc, 100 cycles
13	Electricals (AC, DC, Functional)	Test to datasheet. Read & record data. Tri-temp (-40°C, 25°C, 85°C).	45 pc	22 pc
14	High Temperature Operating Life (HTOL)	MIL-STD-883, Method 1005, 125°C		
15	Electricals (AC, DC, Functional)	Test to datasheet. Read & record data. Tri-temp (-40°C, 25°C, 85°C).		
16	Destructive Physical Analysis (DPA)	2 pc from temp cycled samples, rest from life tested samples	5 pc	3 pc
17	Reliability Characterization	See Table 4	10 pc	
18	Radiation Characterization	TID and SEE per radiation specialist recommendation	Per rad specialist	

[1] 3% PDA for functional parameters at 25°C.

After serializing all components in the lot, initial electrical testing is performed, followed by burn-in and final electricals. Electrical testing involves testing the full array of data sheet parameters plus a set of functional tests. Functional tests usually involve erasing, programming, and reading the memory array with a variety of data patterns and algorithms tailored for the targeted application. For example, if the application were storage of radar data, then serially programming blocks page by page would be appropriate. But if the application were file system storage, then exercising the memory in a more random fashion (skipping pages, partial programming, etc.) would be more appropriate. Operating the device as close to flight-like conditions is important.

Next, delta calculations are usually made on operating and stand-by current parameters. The mission may want to add other critical parameters to the delta criteria, as warranted.

Then a stabilization bake is performed to prepare flight parts for storage. The purpose of this step is mostly to remove any moisture that may have been absorbed by the packaging during exposure to the atmosphere during screening.

After screening is complete, samples shall be chosen at random from the lot for qualification testing. Qualification test samples first undergo preconditioning, a procedure that subjects the samples to thermal conditions meant to replicate the board assembly reflow process. This stresses the devices in a way so as to put them in a state similar to how they would be on a flight board after assembly.

Then temperature cycling, life testing, DPA, reliability characterization, and radiation characterization can be performed in parallel.

Failure Criteria

In Table 3, the steps where failures can occur include electrical tests, delta calculations, and DPA.

Electrical failures mean the device did not perform within datasheet specification, including AC, DC, timing, or functional operation at a given temperature. A delta calculation failure means operational or standby current (or whatever parameters are most important to the mission) shifted greater than 10% over the course of some stressful event such as burn-in, life test, or temperature cycling. If many units failed delta calculation or had electrical measurements out of specification, then the PDA calculation may warrant rejection of the lot. DPA failure means the device did not meet one of the criteria in MIL-STD-1580.

Reliability and radiation characterization could also have unfavorable results, without necessarily having device failures (devices operate within specification but outside mission requirements). For example, bit error rates may be too high during endurance cycling or devices cannot operate beyond 5 krad (Si) total-dose exposure. A failure here may simply mean rejection of the entire lot if performance is not found to be suitable for the application.

3-4 Reliability Characterization

Reliability characterization involves endurance cycling, data retention bakes, and read disturb testing. The goal of this testing is to characterize the bit error rate over the life of the device when operated according to a specific use condition or application. This type of testing is more important for MLC technologies than SLC, although it is recommended for both types whenever bit error rates may be a concern.

All NAND Flash devices will see bit errors (see Section 2-2). The goal of reliability characterization testing is to see if the bit error rate of the device during application-specific usage cases meets mission requirements. Each application has its own set of requirements. Science data bit error requirements are typically more lenient than file system or boot code requirements. A particular NAND device may be suitable for one application but not another. Also, each NAND product has its own bit error characteristics, so space projects must exhaustively characterize the bit

error rates over the life of the device (program/erase cycles) to see if the bit error rate will meet their requirements. This means performing data retention bakes, endurance cycling, read disturb, and program disturb testing according to Table 4.

This testing can also help the project design the level of error detection and correction (EDAC) needed to meet system requirements. Once the raw bit error rate is known from cycling, retention, and disturb errors, an appropriate EDAC system can be designed to lower the bit error rates to acceptable levels for the application [7]. A complex error correction code with large performance overhead (such as the BCH algorithm) may be needed, or a simpler Hamming code implemented in hardware may suffice. Even the simple Hamming code can lower bit error rates by orders of magnitude.

Table 4. Reliability Characterization Tests

Step	Test	Requirements
1	Endurance Cycling	Up to 1.5x application endurance requirement
2	Data Retention Bake	Three temperatures; at 1.5x application endurance requirement
3	Read Disturb	Multiple reads after 0, 0.5x, 1.0x, and 1.5x application endurance requirement
4	Worst Case UBER Calculation	At 1.5x application endurance requirement

Endurance Cycling

As discussed in Section 2, the primary failure mechanism in Flash is repeated programs and erases via FNT. To characterize these types of failures, devices are repeatedly erased, programmed, and read thousands (MLC) or hundreds of thousands (SLC) of times.

Erasing, programming, and reading an entire multi-gigabit NAND device can take a long time. Cycling the entire array up to the endurance specification could take months of test time. In order to keep test schedules manageable, endurance cycling can be done on a sample of blocks instead. The worst-case procedure would be to cycle only one block at a time, erasing, programming, and reading it as quickly as possible up to, or beyond, the endurance specification before moving to the next block. Cycling one block and then moving on to another block would give certain types of defects created in the first block's oxides time to repair.

The data pattern used should be pseudo-random, similar to the type of data seen in application. Also, the way the block is programmed and read should be similar to the intended use-case. If it's serially programmed and reads science data, then that is how the part should be treated. If the target application is file system data, which is typically programmed and read in a random fashion by hitting pages out of order within the block, then that is how the device should be exercised during cycling.

With each read of the device, the number of bit errors should be recorded. It is also useful to record the bit locations as well. This data set can be used to simulate application-like bit errors when evaluating possible EDAC approaches.

Blocks tested should be distributed across the die. And multiple die should be tested; 45 blocks from 3 die (135 total blocks) would be an appropriate endurance test. Flight temperatures, maximum voltage, and maximum frequency should be used. Blocks should be cycled up to 1.5x the application endurance requirement. For example, if an application requires 2,000 cycles, then blocks should be cycled up to 3,000 times for this test, and the bit error rate recorded for each cycle.

Data Retention Bake

The data retention bake measures how well memory cells hold their charge/data. The process involves programming a known data pattern into the device, baking at high temperature, and then reading the device periodically and recording the number of bit errors. The data used is not important and will not have an effect on the activation energy calculation. Typically all 0s is used, meaning all bits are charged and participate in the test. A data retention bake of all 1s in an SLC device would never produce errors - they all begin the test with their floating gates depleted. However, an all 1s data pattern in an MLC device would actually show errors due to the more complicated voltage levels used in these devices. An MLC bit could transition from 0 to 1, back to 0, and finally to 1 during a data retention test.

Typically data retention bakes last until 50% of the bits have failed, ensuring that the experiment has been fully carried into the middle (T_{50}) of the bit population. This produces the most accurate measurements and ultimately the most accurate estimation of the activation energy.

At room temperature, Flash cells can hold their data for many years. In order to accelerate the floating gate leakage process, devices must be baked at high temperature in order to get bits to change state in a timely manner. The data retention bake is done at three temperatures in order to produce an accurate estimation of activation energy [8]. Also, because data retention gets worse with cycling, it is recommended that the data retention bake test be performed on devices that have previously been cycled at least 1.5x application endurance requirement to ensure retention is not an issue at end-of-life.

Read Disturb

The purpose of read disturb testing is to measure how the bit error rate worsens with multiple reads of a block of data. During this test, the block is erased, programmed once, and then read many times. The data pattern, the manner in which the pages are accessed (serial or random), and the number of successive reads should match the intended use-case parameters. The test is then repeated after cycling the device as read disturb gets progressively worse with cycling. Recommended read disturb test points are 0x, 0.5x, 1.0x, and 1.5x application endurance requirement. It is also recommended that flight temperatures, maximum voltage, and maximum frequency are used.

UBER Calculation

The worse-case bit error rate seen during application will be when the effects of cycling and read disturb are combined. This means multiple reads of a data set from a device that has previously undergone many program/erase cycles. It is recommended that the worst-case UBER calculation be performed using the raw BER measured after 1.5x application endurance requirement and 1.5x application read requirement.

This means that if the application intends to cycle devices 2,000 times and could read any set of data up to 50 times between erases, then the worst-case UBER should be calculated using BER after 75 reads of a set of data from a device that was previously cycled 3,000 times. Because BER can vary significantly from cycle to cycle, the BER at "1.5x cycling requirement and 1.5x read requirement" would most accurately be reported as the average BER measured around these limits. Continuing with this example, the most accurate way to report BER at "3,000 cycles and 75 reads" would be to calculate average BER from read cycles 70 thru 80 as measured after 3,000 erase/program/read cycles.

Using this raw worst-case BER, UBER can be calculated using the binomial distribution. There are four conditions a data set must meet in order for the binomial distribution to apply: 1) only two outcomes are possible, 2) there exists a fixed number of trials, 3) there exists a fixed probability of success from trial to trial, and 4) the outcomes are independent from trial to trial.

Probability calculations using the binomial distribution look like this:

$$P(k\ out\ of\ n) = \frac{n!}{k!\,(n-k)!}\left(p^k\right)\left(q^{n-k}\right)$$

where n is the number of trials, k is the number of occurrences of some event, p is probability of k happening in any given trial, and q is the probability it does not happen, or 1-p. In the case of reading a NAND Flash device, n is the number of bits read, k is the number of bit errors, and p is the raw BER. P(k out of n) therefore means "the probability of k bit errors out of n bits read."

For example, if the particular application utilizes five NANDs in parallel to create a 40-bit word, and this 40-bit word is subject to a 2-bit-detection, 1-bit-correction Hamming Code EDAC, then an uncorrectable bit error is seen anytime 2 or more bit errors occur in the 40-bit word. The total UBER is therefore the sum of P(2 out of 40) thru P(40 out of 40). Fortunately, in practice, BERs are so low that the values obtained from P(3 out of 40) and beyond are insignificant and can usually be ignored. If the raw BER is found to be 1e-5 during reliability testing, then P(2 out of 40) is 7.797e-8. The total UBER, P(*2 or more* out of 40), is insignificantly larger at 7.798e-8.

It should be noted that although the process of reading any memory meets the first two binomial distribution criteria perfectly – the bit either passes or fails, and there is always a fixed number of bits being read – the process of reading a NAND Flash does not fully meet the third and fourth criteria. It has been shown that bit errors are not randomly distributed [2], and as this report has described, NAND devices are consumable and each operation has an affect on the reliability of the next operation. However, the distribution of errors has no affect on the *average* bit error rate. So the calculation of average UBER is still accurate, but the end user should be aware that certain pages or blocks may have higher or lower BER/UBER. And although it may seem reading NAND Flash violates the fourth criteria due to endurance-cycling-related degradation and disturb phenomena, it actually does not. When considering the binomial distribution, "trial" means "bit." Each trial has no extrinsic affect on the other bits. Reading a bit is not like pulling cards from a deck of 52; when choosing a card from the deck, and leaving it out for the next trial, that first trial has a clear impact on the outcome of the next (one less possible outcome with the card removed). Plus, when using five NAND devices in parallel, only 8 of the 40 bits are from the same device.

Overall, the binomial distribution fits the process of reading memories very well, and it is standard practice throughout the industry to invoke it whenever making these types of calculations. Fig. 6 shows what the UBER looks like in a 40-bit word after going through a 1-bit correction Hamming Code EDAC. This plot shows that even a simple Hamming Code EDAC can improve BER by orders of magnitude. For a worst-case raw BER of 1e-6, the corresponding UBER is 7.8e-10.

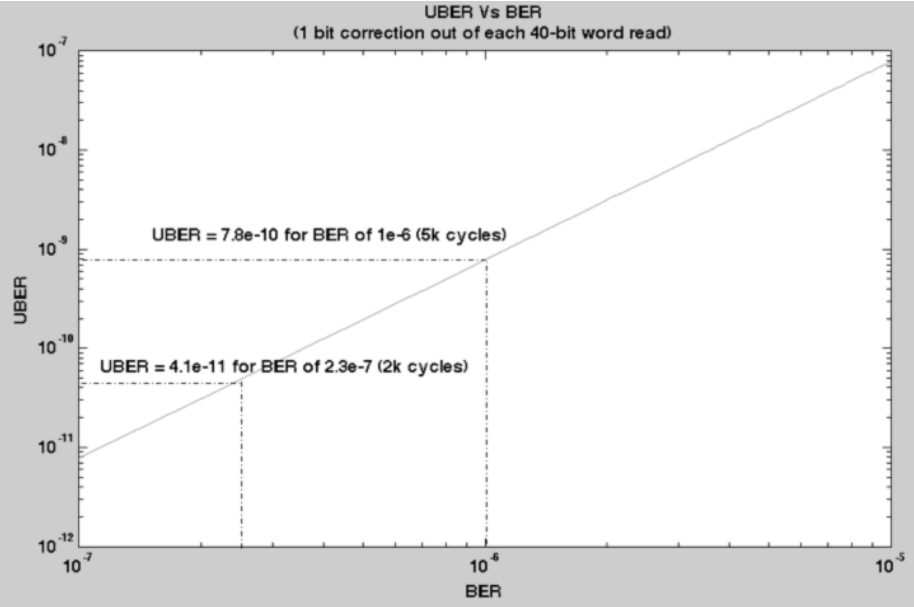

Fig. 6. UBER Vs. BER using binomial distribution, assuming 40 bits read with EDAC capable of correcting 1 out of 40. Two hypothetical data points are indicated: one for a NAND device cycled 2,000 times and one cycled 5,000 times.

3-5 Radiation Characterization

Every flight lot of each device type should be tested for radiation effects. Lot-to-lot variation can prove some lots of the same device meet mission radiation requirements while some do not. NAND Flash should be tested for total dose irradiation (TID), single event upset (SEU), and single event fault interrupt (SEFI).

NAND Flash devices can be tested for TID in two modes: Refresh and No Refresh [9]. In Refresh mode, the device under test (DUT) is erased, programmed, and read at each irradiation step. In No Refresh mode, the DUT is erased and programmed once to start the test, and then only read at each irradiation step. The bit error rates will be much higher in No Refresh mode. The purpose of testing in both modes is to exercise the DUT under two very different use cases.

Refresh Mode:
a. Erase, program, and read.
b. Irradiate DUTs with static bias.
c. Read data.
d. Repeat steps a to c for each radiation increment.

No-Refresh Mode (Read Only):
a. Erase, program, and read.
b. Irradiate DUTs with static bias.
c. Read data.
d. Repeat steps b to c for each radiation increment.

In addition to TID testing, NAND Flash should be tested for SEU and SEFI [8]. Test parameters and conditions should be determined by the project radiation specialist and geared towards that particular application.

Ultimately, UBER calculations should be performed using a worst-case BER from the combined effects of endurance cycling, read disturb, and radiation effects. Radiation test samples may include virgin parts as well as those that have previously undergone endurance cycling. However, in some cases, especially with MLC, the BER from the device without irradiation is large enough so that its effects on BER are insignificant and can be ignored [2].

In general, a rule of thumb is that for SLC, radiation has a real effect on BER, but with MLC, the baseline BER inherent to the device dominates the effects from radiation.

4 – System Level Considerations and Other Recommendations

4-1 Error Detection and Correction (EDAC)

EDAC is always required when using NAND Flash. For SLC devices, simple Hamming codes that can detect 2 errors and correct 1 error per word are usually sufficient. With MLC devices, more sophisticated BCH algorithms may be required [7]. Reviewing the ECC specification on the manufacturer's datasheet can assist in determining what level of EDAC is required. Also, because manufacturers tend to quote worst-case bit error rates under strenuous use cases at end of product life, the reliability characterization outline herein can provide a more accurate picture of the bit error rates expected for a particular use-case and application.

4-2 Wear Leveling

Users of NAND Flash should be careful not to exceed the endurance specification of the part by not erasing or programming any given block too many times. The program/erase stress should be distributed evenly over the part. Many wear-leveling algorithms can be used to accomplish this [10].

4-3 Other Considerations

Part Selection, Pure Tin Leads

When selecting a NAND Flash product for space application, several factors should be considered including encapsulate material, whether it contains corrosive materials, the glass transition temperature of that encapsulate, the package lead finish, and device bit error rate (required ECC) specification.

The material used in the plastic encapsulate should be free of corrosive materials and have leads that are not pure tin. If they are, then they must be retinned or solder-dipped before installation onto flight boards.

Also, before making the final decision on a device, a sample of 3 units should be procured so that the glass transition temperature of the material can be measured. It needs to be high enough to leave margin above the highest expected flight temperature. And all subsequent testing on the device should stay below that temperature.

Finally, the space project must pay close attention to the bit error rate specification in the manufacturer's datasheet. If a 12-bit BCH algorithm is recommended to use the part, the project should plan to have that capability in their design.

Disturb Error Mitigation

To limit disturb related bit errors, the end-user should limit the number of reads between programming, sequentially program pages in a block, and minimize partial-page programming [3].

Moisture Sensitivity, Proper PEM Storage

PEMs have much more variation in raw materials than hermetically sealed packages, which create challenges for successful long-term storage. The encapsulates used are hygroscopic and absorb water.

All PEMs are delivered with a known moisture sensitivity rating. This dictates the amount of time they are allowed outside of the drypack, along with storage temperature and humidity restrictions. PEMs should be handled and stored accordingly.

Many plastic devices are "rated" as non-moisture sensitive, but this is related to their resistance to solder heat/delamination/popcorning, not for long-term storage. The common misunderstanding is that moisture is only a problem when a device is exposed to the heat of solder reflow. In reality, moisture is a problem when combined with long-term leaching of materials in the mold compound, harmful gases, or materials contaminating the exterior of the plastic package, which can result in degradation of product lifetime.

All PEMs should be stored in dry bagging in temperature and humidity controlled rooms.

Electrostatic Discharge (ESD) Sensitivity

Most NAND Flash devices are classified has having ESD sensitivities of Class 1C (>1000 V HBM) or better. Most standard ESD practices (static dissipative wrist straps, workbenches, and storage bags) will prevent damaging the device.

5 – Conclusion

NAND Flash memory provides a very attractive high density, low power, and nonvolatile memory solution. Despite its commercial heritage, this technology can be qualified for a wide range of space applications. Hopefully this guideline provides a screening and qualification blueprint for future space missions that can benefit from this technology.

6 – References

[1] http://www.spacemicro.com/space_div/se_div.htm#flash

[2] J. Heidecker, et al., "Qualification of 128 Gb MLC NAND Flash for SMAP space mission," Integrated Reliability Workshop Final Report (IRW), 2010 IEEE International, October 17-21, 2010.

[3] D. Sheldon, "Disturb Testing in Flash Memories," *NASA Electronics and Packaging Program (NEPP)*, 2009.

[4] M. White, "Scaled CMOS Reliability Users Guide," *NASA Electronics and Packaging Program (NEPP),* 2011.

[5] M. Cooper, "Plastic Encapsulated Microcircuits Reliability/Usage Guidelines for Space Applications," NASA JPL DocID 62212.

[6] R. Menke, "Part Engineering Technical Standard," NASA JPL DocID 78157.

[7] S. Lin, and D. J. Costello, Jr., Error Control Coding: Fundamentals and Applications, 1983.

[8] Trindade and Tobias, "Applied Reliability," 1995.

[9] F. Irom, "Comparison of TID Response and SEE Characterization of Single- and Multi-Level High Density NAND Flash Memories," *RADECS*, 2009.

[10] Wear Leveling Techniques in NAND Flash Devices, Micron, TN-29-42.