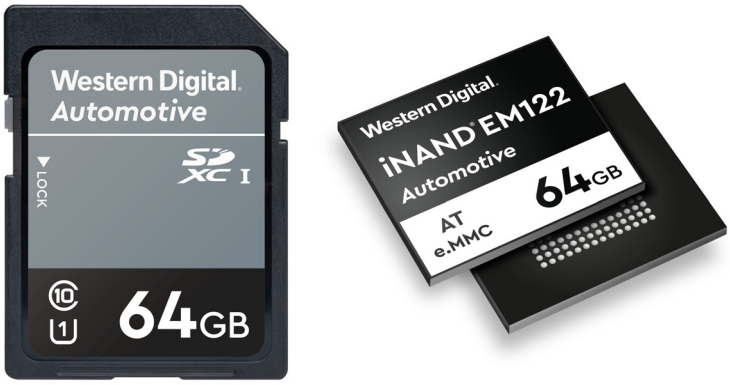
Western Digital.

WHITE PAPER

Flash 101 and Flash Management

A detailed overview of flash and flash management



Western Digital Technologies, Inc. or its affiliates' (collectively "Western Digital") general policy does not recommend the use of its products in life support applications where in a failure or malfunction of the product may directly threaten life or injury. Per Western Digital Terms and Conditions of Sale, the user of Western Digital products in life support applications assumes all risk of such use and indemnifies Western Digital against all damages.

This document is for information use only and is subject to change without prior notice. Western Digital assumes no responsibility for any errors that may appear in this document, nor for incidental or consequential damages resulting from the furnishing, performance or use of this material.

Absent a written agreement signed by Western Digital or its authorized representative to the contrary, Western Digital explicitly disclaims any express and implied warranties and indemnities of any kind that may, or could, be associated with this document and related material, and any user of this document or related material agrees to such disclaimer as a precondition to receipt and usage hereof.

Each user of this document or any product referred to herein expressly waives all guaranties and warranties of any kind associated with this document any related materials or such product, whether expressed or implied, including without limitation, any implied warranty of merchantability or fitness for a particular purpose or non-infringement. Each user of this document or any product referred to herein also expressly agrees Western Digital shall not be liable for any incidental, punitive, indirect, special, or consequential damages, including without limitation physical injury or death, property damage, lost data, loss of profits or costs of procurement of substitute goods, technology, or services, arising out of or related to this document, any related materials or any product referred to herein, regardless of whether such damages are based on tort, warranty, contract, or any other legal theory, even if advised of the possibility of such damages.

This document and its contents, including diagrams, schematics, methodology, work product, and intellectual property rights described in, associated with, or implied by this document, are the sole and exclusive property of Western Digital. No intellectual property license, express or implied, is granted by Western Digital associated with the document recipient's receipt, access and/or use of this document or the products referred to herein; Western Digital retains all rights hereto.

This document and Western Digital communications to the user associated therewith, shall be treated as Western Digital Corporation's proprietary and confidential information, protected by the recipient as such, and used by the recipient only for the purpose authorized in writing by Western Digital Corporation. This document shall be covered as Western Digital Corporation's confidential information under all applicable nondisclosure agreements between the recipient and Western Digital Corporation.

Table of Contents

1.0 Introduction

The inherent nature of NAND flash technology requires sophisticated flash management techniques to make it a practical storage medium for computing systems. Challenges intrinsic to using NAND flash in a managed device (e.MMC, SD, micro SD etc.) include:

- Requirement to erase before program (writing)
- Wear out mechanism that limits service life
- Data errors caused by write and read disturb
- Data retention errors
- Management of initial and runtime bad blocks

With proper flash management techniques, these characteristics of NAND flash can be managed to provide a highly reliable data storage device.

Five significant factors influencing reliability, performance, and write endurance of the managed NAND devices are:

- Use of Single-Level Cell (SLC) vs. Multi-Level Cell (MLC) NAND flash technology
- Wear-leveling algorithms
- Ensuring data integrity through Bad Block management techniques
- Use of error detection and correction techniques
- Write amplification

Proper implementation of sophisticated flash management techniques will deliver products with high reliability, long service life, high performance, and excellent data integrity characteristics.

This white paper provides an overview of NAND flash technology, its intrinsic characteristics, and explains how proper flash management techniques address specific NAND issues to create reliable managed NAND device with a long service life.

2.0 NAND Flash Technology

NAND flash is a nonvolatile solid state memory with the capability to retain stored data when unpowered. NAND and NOR are the two fundamental flash architectures used in electronic systems today. Both NOR and NAND Flash memory were invented by Dr. Fujio Masuoka in 1984 [1]. The first commercial NAND flash chip was introduced in 1989.

NAND flash offers faster erase and write time up to ten times compared to NOR flash [2]. It requires a smaller chip area per cell (compared to NOR), thus allowing greater storage densities and lower cost per bit. NAND flash achieves these advantages by sharing some of the common areas of the storage transistor through strings of serially connected transistors. NOR devices require additional control circuits to independently access each storage transistor for random, independent addressability.

NAND flash access is similar to other block-oriented storage devices such as hard disks and optical media, and therefore is frequently used in mass-storage devices such as memory cards, e.MMC devices and USB flash drives.

e.MMC protocol implemented hardware boot schema, SnD (Store and Download) allowing to replace the NOR as a boot device with limited rom changes on the host size. For more details, refer to the e.MMC JEDEC specification (membership required).

Today, two NAND flash technologies, SLC (Single-Level Cell) and MLC (Multi-Level Cell) that can have 2 or 3-bit per cell, service different applications. Section 3 explains in detail the differences between these two technologies.

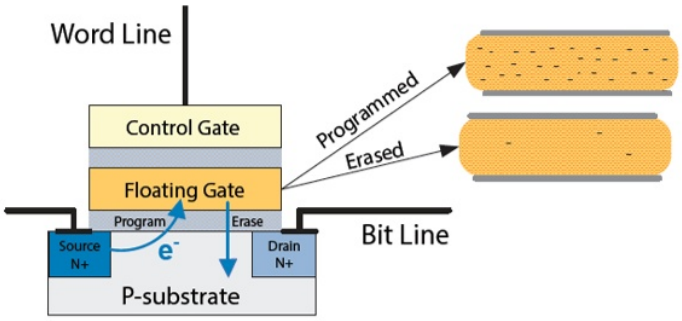
3.0 NAND Flash Cell

The basic NAND flash cell is a floating gate transistor with the bit value determined by the amount of charge trapped in the floating gate. NAND flash uses tunnel injection for writing/programming and tunnel release for erasing the cell[1].

In SLC (Single-Level Cell):

- Writing (e.g., programming) to a cell causes the accumulation of negative charge in the floating gate, resulting in a "0" bit value for that cell.
- Erasing a cell removes the negative charge in the floating gate, resulting in a "1" bit value for that cell. To change the bit content of a cell from "0" to "1", the cell must be erased. Due to the NAND architecture of sharing bit control lines and word control lines across multiple storage transistors, erasing a cell requires erasing the entire Erase Block, which contains that cell.

Figure 3-1. NAND Flash Cell Architecture[1]



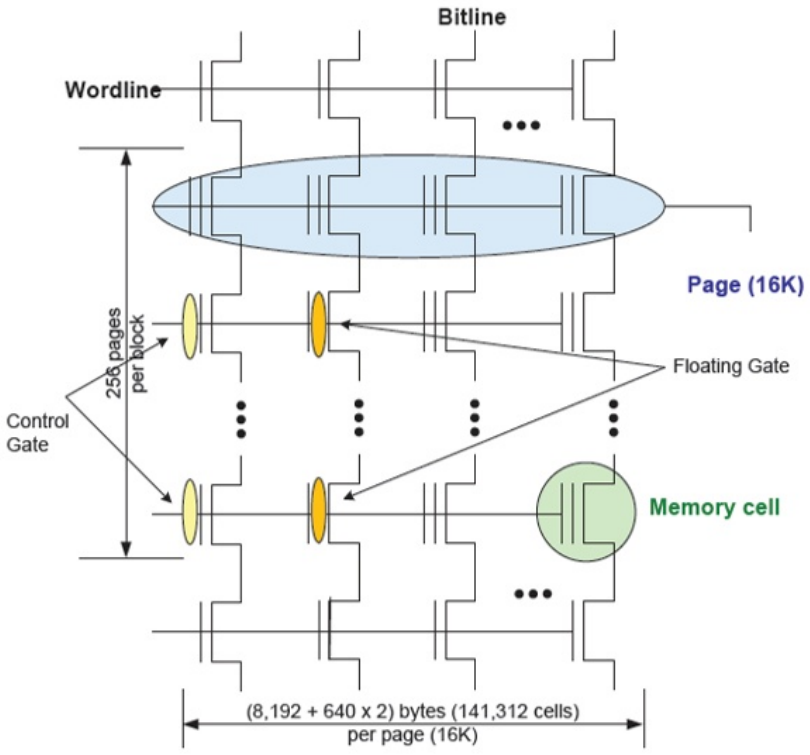
4.0 NAND Flash Architecture

NAND flash memory stores the information in an array of floating-gate transistors, i.e. memory cells, combined into bit and word lines. The serial cell architecture of NAND explains the device name. NAND (Not AND) is the Boolean logic reference to how information is read out of these cells. Each single level cell (SLC) transistor stores one bit of data, and each multi-level cell (MLC) NAND stores multiple bits of data in each cell. Figure 4-1 below shows the NAND flash architecture of a 64Gb MLC19nm flash[2], where 17,664 cells are located on the same wordline to create a 16KB pages. An Erase Block consists of 256 pages, every two pages (in 2 bits per cell) occupy single wordline over 17,664 bitlines.

[1] Jitu J. Makwana and Dr. Dieter K. Schroder, A Non-Volatile Memory Overview, http://aplawrence.com/Makwana/nonvolmem.html.

[2] Jim Cooke, Inconvenient Truths of NAND Flash, Flash Memory Summit, August 2007.

Figure 4-1. NAND Flash Architecture (64GB MLC 19nm)[1]



4.1 Erase Blocks and Pages

A page is the smallest area of the flash memory that supports a write operation and consists of all the memory cells on the same wordline. An Erase Block is the smallest area of the flash memory that can be erased in a single operation. Page and block sizes differ per manufacturer and flash generation.

For example:

19nm 64Gb MLC flash contains 16KB page size and 4MB block size, as shown in Figure 4-1 above. 16KB page size corresponds to 16,384 bytes that are dedicated for data and 1,280 bytes that are available for control and Error Correction Code (ECC) information.

4.2 SLC vs. MLC NAND Technology

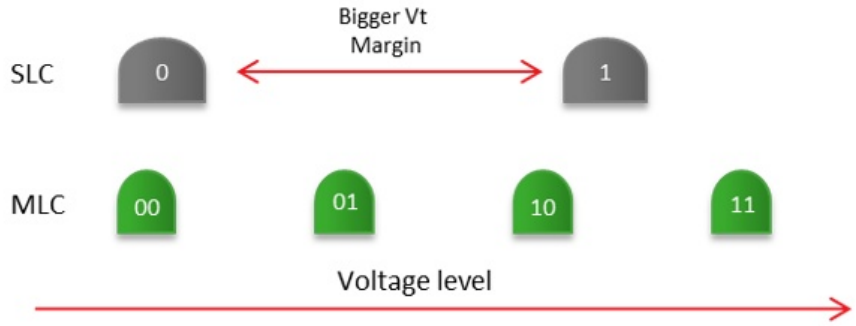
Multi-Level Cell (MLC) NAND and Single-Level Cell (SLC) NAND offer capabilities that serve two very different classes of applications – those requiring the lowest cost-per-bit, and those demanding higher endurance.

MLC NAND flash allows each memory cell to store multiple bits of information, compared to the one bit per cell for SLC NAND flash. As a result, MLC NAND offers a larger capacity, twice the density of SLC, and at a cost and reliability point targeted for consumer products such as cell phones, digital cameras, USB drives, and memory cards.

[1] Jim Cooke, Inconvenient Truths of NAND Flash, Flash Memory Summit, August 2007.

SLC NAND provides faster write speed and longer write endurance, making it popular for use with applications that require high performance like enterprise level SSDs and high endurance applications like Time Shift Buffer that constantly write to the device and needs viability of multi-year service life.

Figure 4-2. SLC vs. 2-bit per cell MLC NAND Technology



As shown in Figure 4-2 above, both SLC- and MLC-based devices use about the same size voltage window. These separations between adjacent voltage levels called Vt (Voltage thresholds) are smaller in MLC technology. Smaller distance between Vt impacts:

- Write performance: since charging the cells to the correct voltage levels requires more iterations.
- NAND write endurance: since more iterations means more stress on the physical cell.

Western Digital e.MMC line (Managed NAND) takes advantage of these basic NAND features to introduce ultrafast write performance with its SmartSLC™ feature. Performance and endurance are product dependent and varies based on memory generation and system architecture.

NAND industry is scaling down wafer lithography every year to reduce cost. This process inevitably imposes reduction in NAND reliability that will likely result in lower endurance and performance capabilities. When choosing your NAND supplier make sure to understand the performance and endurance of your current NAND product and requirements.

5.0 Inherent NAND FLASH Challenges

NAND flash could potentially suffer from reliability, write endurance, and data retention problems that require complex management solutions. Due to the increased density with each new flash generation, these problems are increasingly becoming more apparent, requiring more and more sophisticated NAND flash controllers and system solutions. Some intrinsic limitations include:

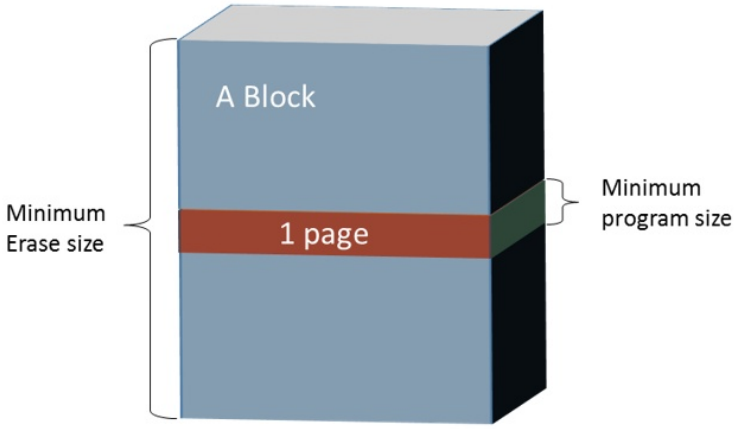
- Need to erase before writing
- Wear out mechanism that limits service life
- Data errors caused by write and read disturb
- Data retention errors
- Management of initial and runtime bad blocks

5.1 Erase before Write

NAND access patterns are:

- Read and Write in pages
- Erase in blocks

Figure 5-1. Page Write vs. Block Erase



Blocks are consecutive groups of pages. Page and block size depend on the lithography and the NAND manufacture. Before pages can be written, the block in which the page belong to must be erased.

The Flash management enables transparent logical access to the NAND flash. However, it is important to enable as good as possible NAND friendly access pattern to limit the "overhead". This "overhead" or "housekeeping" activities that are triggered by the logical to physical management will affect the following:

- Write Amplification
- Product Performance

Customers are encouraged to optimize the load especially in write intensive applications to improve product longevity. More information on system optimization can be found in *Section 7.0 System Optimization Techniques on page 14*.

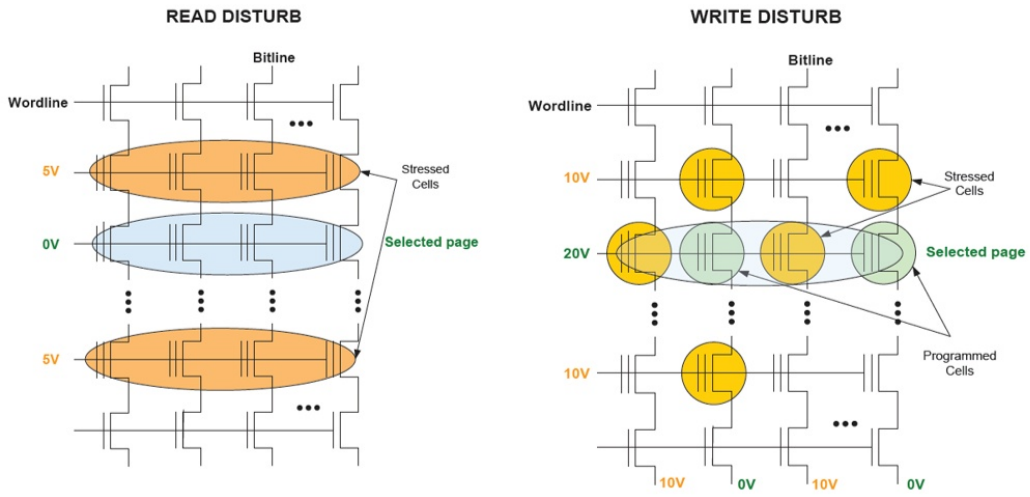
5.2 Read/Write Disturb

NAND flash is prone to bit flips - cells that are not meant to be accessed during a specific read or write operation can change contents due to read and write activities in adjacent cells or pages.

- Read Disturb: A read disturb occurs when a cell that is not being read receives elevated voltage stress. Stressed cells are always in the block that is being read and are always on a page that is not being read. The probability of read disturb is much lower than is a write disturb.
- Write Disturb: A write disturb occurs when a cell that is not being programmed receives elevated voltage stress. Stressed cells are always in the block that is being programmed and can be either on the page that is programmed (but cell was not selected), or on any page within the same block.

Erasing the cell resets the cell to its original state, eliminating the data and, consequently, the data errors which resulted from the read or write disturbs. An ECC mechanism in the data flow path detects bit flips and corrects them before providing the data to the host. As flash cell geometries decrease and more cells are placed onto wafers, the probability of errors and bit flips increases and NAND flash controllers require more powerful error detection/correction (EDC/ECC) algorithms.

Figure 5-2. Read/Write Disturb NAND Flash[1]



5.3 Data Retention Errors

Data retention defines how long the written data remains valid within a memory device. The data retention period has correlation to the cell wear status. Fresh cell (cell that was not programmed and erased many times) will keep the data longer than a worn out cell. Temperature is a another factor in the formula to define the period of time of data retention, the higher the temperature the shorter the time the data will be kept.

JEDEC specification for NAND is 1 year at 55^C for EOL (End of Life) device and 10 years at 55^C for fresh device. Some product lines may offer various configurations of data retention/temperature/endurance variables.

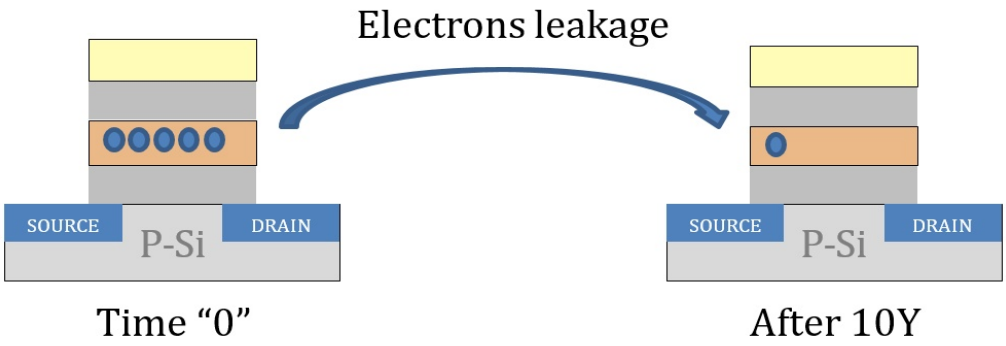
The limitation of data retention of NAND devices derives from a charge leakage from the floating gate, which tends to slowly change the cell's voltage level from its initial level to a different level, as shown in Figure 5-3 below. The new level may be incorrectly interpreted as a different logical value which results in data errors.

The data retention time is inversely related to the number of Write/Erase cycles, which means the blocks that have been erased multiple times will have a shorter data retention life than blocks with lower Write/Erase cycles.

To overcome the predicted bit error rate, an appropriate ECC mechanism needs to be implemented to detect data retention errors and correct them before providing the data to the host.

[1] Jim Cooke, Inconvenient Truths of NAND Flash, Flash Memory Summit, August 2007.

Figure 5-3. Data retention errors through charge leakage[1]



5.4 Bad Blocks

There are two types of bad blocks in a NAND flash device:

- Initial Bad Blocks: Due to production yield constraints and the pressure to keep costs low, NAND flash devices ship from the factory with a number of bad blocks.
- Accumulated/Dynamic Bad Blocks: Due to multiple write/erase cycles, trapped electrons in the dielectric cause a permanent shift in the voltage levels of the cells. When the voltage level shifts enough this will be observed as a read, write, or erase failure.

Bad Block management is required to map out both the initial Bad Blocks, as well Bad Blocks that were accumulated during device operation, while guaranteeing the initial user capacity.

Monitoring Dynamic Bad Blocks cannot guaranty interpretation of the NAND health level. For a comprehensive health level, consult the device SMART report that can provide details about the utilization of the endurance and other diagnostics on the NAND.

5.5 Limited Number of Writes

NAND flash memory has a finite number of Program/Erase (P/E) cycles for two reasons:

- Electrons that are trapped (i.e. trap-up) in the thin oxide layer that insulates the floating gate;
- Break down of the oxide structure, due to hot carrier injection[2].

Once the damage to the oxide layer is large enough, it becomes increasingly difficult for electrons to travel between the P-substrate and the floating gate, as shown in Figure 5-4 below. The Erase Block encompassing the oxide layer cannot be erased properly with the standard threshold voltages thus needs to be retired and added to the pool of Bad Blocks.

---

1 Richard W. Hamming, Error Detecting and Error Correcting Codes, Bell System Technical Journal 26(2):147-160, 1950.

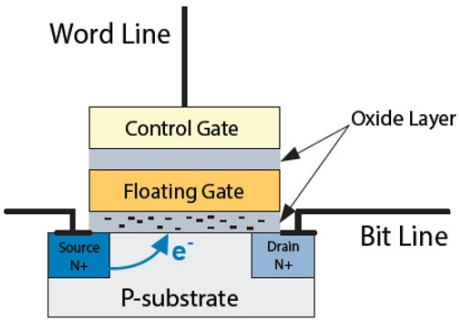
2 Jitu J. Makwana and Dr. Dieter K. Schroder, A Non-Volatile Memory Overview, http://aplawrence.com/Makwana/nonvolmem.html.

The number of Program /Erase (P/E) cycles that NAND flash manufacturers specify for their flash devices is an indication of the expected wear out of the oxide layer.

With each process shrink, flash manufacturers are facing challenges to maintain the same number of write/erase cycles and increase the ECC requirements for the flash.

Flash management techniques, such as Wear Leveling, Error Correction, and Bad Block management are required to overcome and manage the flash wear out limitation.

Figure 5-4. Flash wear out - electrons cannot pass the oxide layer



6.0 Flash Management Techniques

Proper flash management techniques must incorporate mechanisms to overcome the limitations inherent to NAND Flash. A combination of hardware and software solutions is used to manage and overcome the NAND flash limitations. Host optimization and use-case analysis can help further improve the reliability of the system and longevity of the storage.

Flash Management mission is to create a Logical to Physical layer that is transparent to the host and provide logical read and write services. Along with that the Flash management needs to provide reliable product and handle the NAND intrinsic limitations making it transparent to the host. This section will elaborate on the hardware and software mechanisms that are designed to improve product reliability.

6.1 Wear Leveling

Wear Leveling ensures even distribution of erase operations on all blocks within the NAND flash. That is, each block within the NAND flash is erased and written approximately the same number of times as every other block within the device.

To understand Wear Leveling, one needs to understand the different addressing schemes in a system. The operating system (OS) uses Logical Block Addressing (LBA) to read and write to the device, the flash controller uses physical addresses on the flash to read and write data.

Wear Leveling is based on two mechanisms:

- The managed NAND controller has the ability to map an LBA address to different physical locations on the flash. The controller uses a mapping table to keep track of the relationship between the logical block and the physical address.
- Hot count and other parameters of each of the physical blocks are monitored. Once those indicators cross a pre-defined threshold, the wear leveling algorithm will rotate the data to blocks that did not reach this threshold.

This block rotating technique ensures even wear of memory blocks across the flash device. The Wear Leveling process is transparent to the operating system.

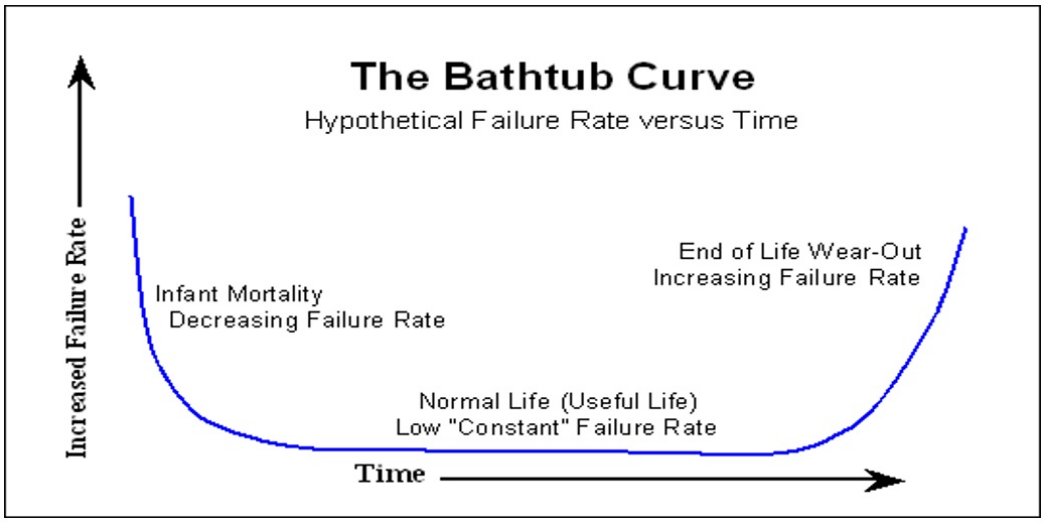
6.2 Exception Management

6.2.1 NAND Failure Management

Product manufacturing involves abundance of testing both of the system level as well as on the NAND die to minimize defect both during the entire product lifetime, but more importantly, during the normal usage lifetime. Illustrated by the "Bathtub" curve graph below, the most ideal is to deliver product to customers during the low failure rate period to maximize customer satisfaction.

Even though the NAND is "cleaned", some of the defects can appear during the life of the product for various reasons. Western Digital firmware is designed to manage defects and protect the user data with special reliability algorithms. The initial goal of meeting automotive requirement of having exceptionally low Defective Parts per Million (DPPM) has over time enabled Western Digital to achieve very low DDPM targets across all Western Digital products.

Figure 6-1. Bathtub reliability curve

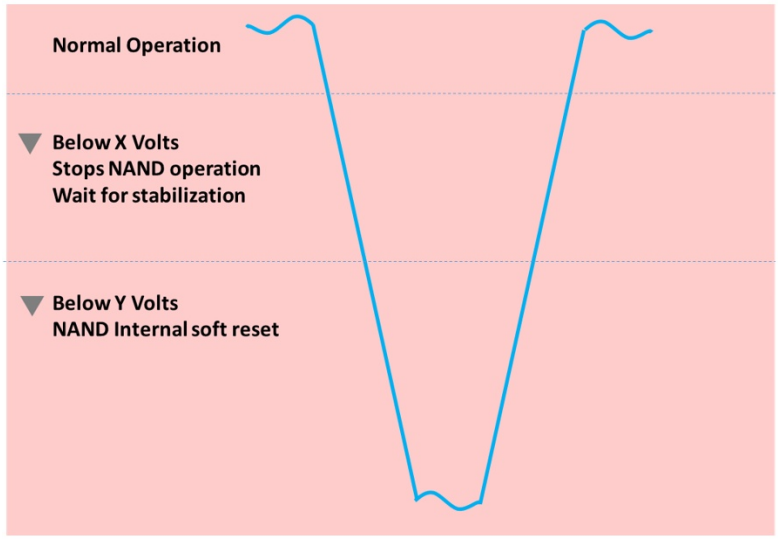


6.2.2 Unstable Power Supply

NAND program operation is sensitive to the fluctuation of the power supply. In cases where power is not stable during a write command, (NAND program) the actual NAND block reach unknown state (not erased nor programed) that may impact the device ability to initialize.

Western Digital e.MMC controller for managed NAND flash includes a voltage detection mechanism to prevent NAND from being programmed in noisy environment. It will also detect the faulty program operation to protect the product. In the graph below you can find an example of a product behavior during different voltage thresholds. This behavior is a design decision that may vary on different products.

Figure 6-2. e.MMC voltage threshold behavior example



6.3 Error Detection and Correction

One of the key factors to increase flash reliability as well as write endurance is the implementation of an Error Detection and Correction mechanism. The three most popular Error Correction algorithms used with NAND flash technology today are:

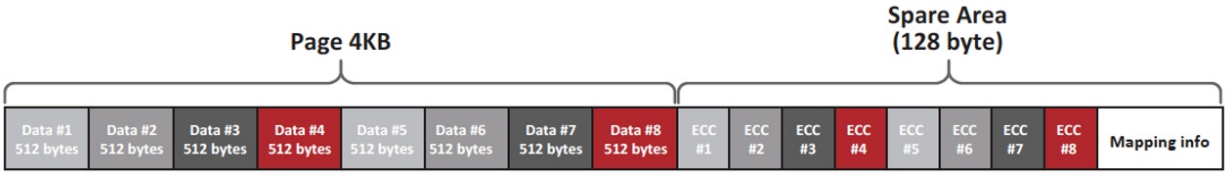
- BCH (Bose, Ray-Chaudhuri, Hocquenghem): BCH codes were invented in 1959 by Hocquenghem, and independently in 1960 by Bose and Ray Chaudhuri.
- LDPC (Low Density Parity Codes) invented by Gallagher in 1961.

The maximum number of errors that can be detected and corrected is determined by the algorithm. No matter what algorithm is used within the ECC engine of a NAND flash controller, the method of error detection and correction remains the same:

1. Every time a page of data is written to the flash, data is passed through the controller's ECC Engine to create a unique ECC signature.
2. The data and ECC signature are stored together on the flash; the data in the page area, the ECC signature in the spare area.
3. When reading the data back, both data and stored ECC signature are read into the controller. A new ECC signature is generated, based upon the read-back data.
4. The newly created ECC signature is then compared to the original stored ECC signature. If both signatures are the same, no errors have occurred, and the data will be provided to the host. If the two signatures differ, the data will be corrected by the controller before being provided to the host.

Some flash controllers will write the corrected data back to the flash to optimize reliability, while others will not, since there is no guarantee that the data will not show errors again in the future. For 4KB page flash, typically 8 ECC signatures are created when writing data to the flash; one for each 512 bytes of data, as shown in Figure 6-3.

Figure 6-3. Four ECC signatures for 4KB page



Implementing an ECC mechanism improves the overall reliability of the flash device, as read, write and data retention errors are caught and corrected. Less known is the fact that a strong ECC engine is one of the most important factors to increase the life span of a flash device. When blocks start to age, more and more errors will occur on that block. When the ECC engine is not able to correct these errors, a hard "ECC" error occurs and the block will be retired. The more powerful the ECC engine, the more "life" can be squeezed out of a block (even though it shows increasing failures) and the longer the overall lifespan of the flash device.

6.3.1 LDPC vs. BCH

LDPC algorithm is more complex on implementation by comparison. However, it provides up to twice the error correction capability that directly translates to the product longevity (life span) and reliability.

Table 6-1. LDPC vs. BCH

Characteristic	BCH	LDPC
Product Endurance	x	Up to 2*X
Decoding Methods	One Read from the NAND	Iterative Method, can do multiple reads to obtain data when error at suspect.
Decoding Methods Typical Block size	Hard Decision based on Voltage threshold.	Soft Decision based on majority and probability for failure.

7.0 System Optimization Techniques

7.1 Write Amplification

When the host writes a page or blocks to the memory device – internal operations inside the managed NAND device will cause additional writes to occur to the memory (management overhead). Write amplification is the amount of data written to the flash vs. the amount of data written by the Host as shown in the equation below.

$$\frac{\text{Data Written to the Flash}}{\text{Data Written by the Host}} = \text{Write Application (WA)}$$

The higher the write amplification, the faster the device will wear out. Write amplification is directly correlated to the workload. This means pure sequential workload will produce the lowest WA factor and random activity will result in higher WA.

A commonly used metric to describe endurance is Terabytes Written (TBW), which is used to describe how much data can be written over the lifetime of the device, regardless of the warranty period.

7.2 Understand and Optimize Your System

In order to optimize device workload, Western Digital recommends the following actions:

1. **Understand the workload requirements**

   For example, in Pause Live TV (PLTV) buffer use case you should consider the following:

   a. What is the bit-rate streamed to the device?

   b. How many hours per day the device will utilize PLTV feature?

   c. What is the rewind time required?

   d. What is the product longevity requirements?

   e. What is the WA factor?

   From your answers, you can calculate the product / buffer requirements

   a. Product / Buffer size is: (a)*(c)

   b. Product TBW is: (a)*(b)*(d)*(e)

2. **Understand your WA factor**

   Western Digital FAEs can provide instructions on how to collect traces of the workload and analyze them to provide estimated WA factor and recommendations for system optimization.

3. **Optimize your write patterns**

   Generally speaking NAND friendly writes are sequential and in big chunks, as a practice it is recommended to operate in:

   a. Write command address alignment to a minimum of 4KB, preferred 128KB.

   b. Write command length at a minimum of 4KB (for random), preferred 512KB.

8.0 Product Name Change

This section describes the product name change from SanDisk to Western Digital.

Description of Change

- Update company logo to Western Digital in collateral and marketing materials. No impact on product marking.
- Retain iNAND as the product brand for embedded flash drives.
- Assign each letter of the new taxonomy with its unique identification.

Figure 8-1. SanDisk to Western Digital

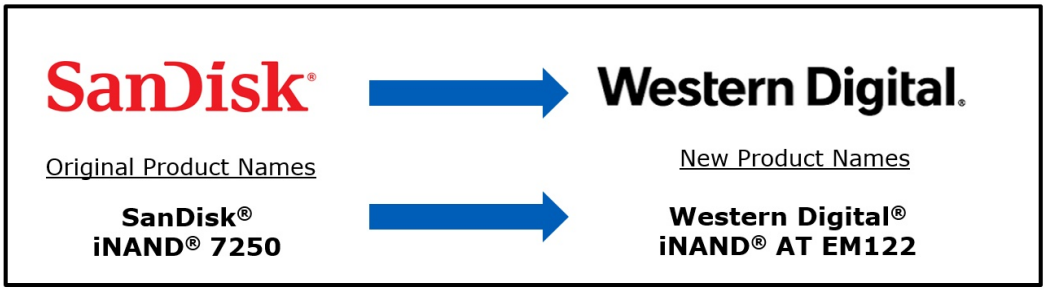
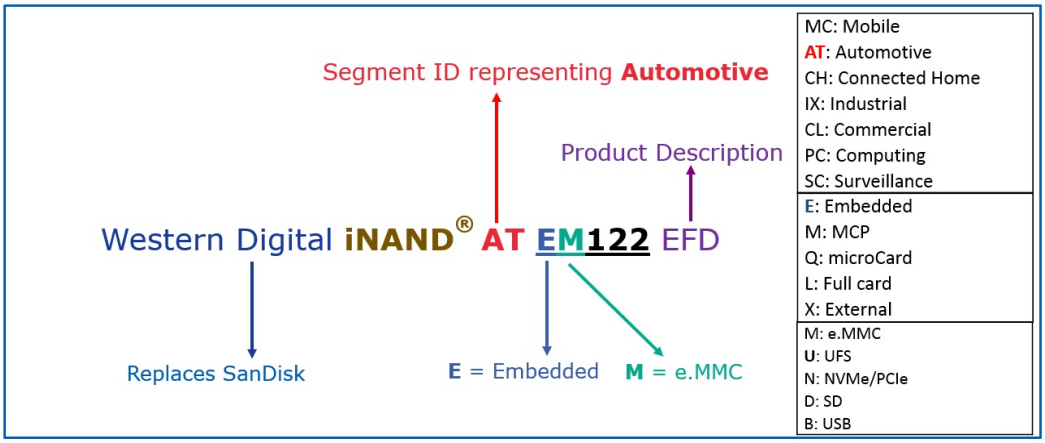


Figure 8-2. Decoding New Naming Structure



Important Exceptions

- No impact on currently shipping or qualified products
- No change to part number/SKU systems
- SanDisk, WD, and HGST part numbers and ordering system are unchanged

If you have any questions or require further assistance, contact oemproducts@wdc.com for support.

9.0 Contact Information

Western Digital Technologies, Inc.
951 SanDisk Drive, Milpitas, CA 95035-7933
Phone: +1-408-801-1000
Fax: +1-408-801-8657
Email: oemproducts@wdc.com

Western Digital®

Western Digital
5601 Great Oaks Parkway
San Jose, CA 95119
U.S.A.

For service and literature:
support.wdc.com
www.wdc.com
800.ASK.4WDC North America
+800.6008.6008 Asia Pacific
00800.ASK.4WDEU Europe (toll free where available)
+31.88.0062100 Europe/Middle East/Africa

August 2018