Flash 101 and Flash Management

A detailed overview of flash and flash management techniques
Flash 101 and Flash Management

A detailed overview of flash and flash management techniques

Table of Contents

1. Introduction ................................................. 4
2. NAND Flash Technology ..................................... 4
3. NAND Flash Cell ............................................. 5
4. NAND Flash Architecture ................................. 5
   4.1 Erase Blocks and Pages ............................ 6
   4.2 SLC and MLC NAND Technology ............... 6
5. Inherent NAND FLASH Challenges ..................... 8
   5.1 Erase Before Write ................................. 8
   5.2 Read/Write Disturbs .............................. 9
   5.3 Data Retention Errors .......................... 9
   5.4 Bad Blocks ........................................ 10
   5.5 Limited Number of Writes ....................... 10
6. Flash Management Techniques ........................ 11
   6.1 Wear Leveling ....................................... 11
   6.2 Exception Management .............................. 12
      6.2.1 NAND Failure Management ................. 12
   6.2.2 Unstable Power Supply ....................... 12
   6.3 Error Detection and Correction .................. 13
      6.3.1 LDPC vs. BCH ............................... 14
7. System Optimization Techniques ....................... 14
   7.1 Write Amplification ................................. 14
   7.2 Understand and Optimize Your System ......... 15
8. References .................................................. 16
1. 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, USB etc.) include:

- Requirement to erase before programming (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

Implementation of sophisticated flash management techniques that are properly implemented 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. 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. Fujiio Masuoka in 1984 [1]. The first commercial NAND flash chip was introduced in 1989.

NAND flash offers faster erase and write times and up to ten times the write endurance 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 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 please refer to the e.MMC JEDEC specification.

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. 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 [3].

In SLC (Single-Level Cell):

- Writing (i.e. 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.

![NAND flash cell architecture](image)

4. 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 2 below shows the NAND flash architecture of a 64Gb MLC19nm flash [4], 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 a single wordline over 17,664 bitlines.
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 2 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 and 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 a cost and reliability grade which perfectly meets consumer products such as cell phones, digital cameras, USB drives, and memory cards.
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.

As shown in Figure 3 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
- Write endurance: since more iterations means more stress on the physical cell

SanDisk e.MMC line (Managed NAND) take advantage of these basic NAND features to introduce ultrafast write performance with its SmartSLC feature as part of an MLC based product, thus offering the best of both worlds — price/capacity of MLC and high write performance of SLC.

Below you can find SanDisk 1Znm memory write performance / endurance details:

<table>
<thead>
<tr>
<th></th>
<th>SLC (1-bit per cell)</th>
<th>MLC (2-bits per cell)</th>
<th>e.MLC (2-bits per cell)</th>
<th>TLC (3-bits per cell)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Write Endurance [Program/Erase Cycles]</td>
<td>30,000 - 50,000</td>
<td>Up to 3,000</td>
<td>Up to 10,000</td>
<td>Up to 1,500</td>
</tr>
<tr>
<td>Write Performance [Per NAND Die]</td>
<td>Up to 150MB/s</td>
<td>Up to 50MB/s</td>
<td>Up to 40MB/s</td>
<td>Up to 15MB/s</td>
</tr>
</tbody>
</table>

The NAND industry has been traditionally scaling down wafer lithography every 12 to 18 months to reduce cost, this process may impose reduction in the raw NAND reliability, performance and endurance capabilities. NAND management plays a critical role in keeping the final product specifications on target of the addressable markets. When choosing your NAND supplier make sure to understand the performance and endurance of current NAND solutions.
5. Inherent NAND FLASH Challenges

Perceived challenges over NAND flash reliability, write endurance, and data retention require complex management solutions. With the increased density for each new flash generation, these challenges are increasingly becoming more apparent, requiring even more sophisticated NAND flash controllers and system solutions to meet the expectation with the desirable performance. Therefore, it is equally important to learn about the intrinsic limitations of NAND Flash:

- 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 4: Page Write vs. Block Erase](image)

Each page can be at a size between 8KB and up to 32KB. Blocks are consecutive groups of pages in the sizes of 4MB up to 8MB depending on the lithography and the NAND manufacturer. Before pages can be written, the block in which the page belongs 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 impacts:

- Write Amplification
- Product Performance

SanDisk supports the eco-system and customers to optimize the load especially in write intensive applications to improve product longevity. More information about how to optimize your system in section 7.
5.2 Read/Write Disturbs

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 disturbances. 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: Read/write-disturbed NAND flash](image-url)

5.3 Data Retention Errors

Data retention defines how long the written data remains valid within a storage 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 for longer time than a worn out cell. Temperature is 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 55c for EOL (End Of Life) and 10 years at 55c for fresh device.

SanDisk High Endurance line supports application specific data retention periods correlated with the endurance, for example 5 days at 75c for temporary files before they move from the client application to the cloud.
The limitation of data retention of NAND devices derives from a charge leakage from the floating gate, called and tends to slowly change the cell’s voltage level from its initial level to a different level, as shown in Figure 6 below. This new level may incorrectly be interpreted as a different logical value.

The data retention time is inversely related to the number of Write/Erase cycles, which means that blocks that have been erased many times 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.

![Figure 6: Data retention errors through charge leakage](image)

### 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 understanding of the NAND health level, for detailed health level please consult the device SMART report that can provide details about the utilization of the endurance and other diagnostics of the NAND.

### 5.5 Limited Number of Writes

NAND flash memory has a finite number of Program/Erase (P/E) cycles, caused by 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[3].
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 7 below. The Erase Block encompassing the oxide layer cannot be erased properly with the standard threshold voltages and needs to be retired and added to the pool of Bad Blocks.

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 7: Flash wear out - electrons cannot pass the oxide layer](image)

6. 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 make it transparent to the host. This section elaborate on the hardware and software mechanisms that improves product reliability.

6.1 Wear Leveling

Wear Leveling helps ensure 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 upon 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 helps ensure 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

A product manufacturing involves abundance of testing both of the system level and on the NAND die to ensure low defect in the product lifetime, ensuring that the product will be delivered to the customer at the low failure rate portion of the “Bathtub” curve graph.

Even though the NAND is “cleaned”, some of the defects can appear during the life of the product. SanDisk firmware is built to manage those defects and protect the user data if possible with special reliability algorithms, this helps SanDisk achieve very low DPPM targets driven mainly by the automotive market, algorithms that are being applied across all SanDisk products.

#### 6.2.2 Unstable Power Supply

NAND program operation is sensitive for 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 programmed) that may impact the device ability to initialize.

SanDisk e.MMC controller includes a voltage detection mechanism to help prevent programing the NAND in noisy environment and detect the faulty program operation to protect the product. In the graph below you can find the e.MMC behavior during different voltage thresholds.

![The Bathtub Curve](image-url)
6.3 Error Detection and Correction

One of the key factors to increase flash reliability and write endurance is the implementation of an Error Detection and Correction mechanism. The two most popular Error Correction algorithms that are 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[6].
- LDPC (Low Density Parity Codes) invented by Gallagher in 1961.
- Hamming: Invented in 1950, named after Richard Hamming[7]. This algorithm can detect and correct single-bit errors, and can detect (but not correct) double-bit errors. This code is mainly used for SLC NAND Flash.

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 is corrected by the controller before being provided to the host.
Some flash controllers will write the corrected data back to the flash media 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 10 below.

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 detect these errors, a hard “ECC” error occurs and the block is 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 drive.

### 6.3.1 LDPC vs. BCH

LDPC algorithm is more complex on implementation however it provides up to twice the error correction capability\(^5\) that directly translates to the product longevity (life span) and reliability.

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>BCH</th>
<th>LDPC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Product Endurance</td>
<td>X</td>
<td>Up to 2(^x)</td>
</tr>
<tr>
<td>Decoding Methods</td>
<td>One Read from the NAND</td>
<td>Iterative Method, can perform multiple reads to obtain data when error is suspected</td>
</tr>
<tr>
<td>Typical Block size</td>
<td>Hard decision based on voltage threshold</td>
<td>Soft decision based on majority and probability of failure</td>
</tr>
</tbody>
</table>

### 7. System Optimization Techniques

#### 7.1 Write Amplification

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.

\[
\text{data written to the flash} / \text{data written by the host} = \text{Write Amplification (WA)}
\]
The higher the write amplification, the quicker the device will wear out. Write amplification is directly correlated to the workload; pure sequential workload will produce the lowest WA factor, and random activity will result in high 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 life of the device.

7.2 Understand and Optimize Your System

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

1. Understand the workload requirements

   For example, in video recording applications, where there is a need to buffer (keep some data) as well as be able to watch (rewind) — answer the below questions:
   
   a. What is the bit-rate streamed to the device?
   b. How many hours per day the device will record?
   c. What is the rewind time required?
   d. What is the product longevity requirements?
   e. What is the WA factor?

   Following the answers of the above you can calculate the estimated product / buffer requirements:
   
   f. Product buffer size = (a)*(c)
   g. Product Terabytes Written (TBW) requirements = (a)*(b)*(d)*(e)

2. Understand your WA factor

   SanDisk FAEs can help 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. References

[5] Based on SanDisk internal analysis

Products, samples and prototypes are subject to update and change for technological and manufacturing purposes. Western Digital Technologies, Inc.’s (“WDT”) general policy does not recommend the use of its products in life support applications wherein a failure or malfunction of the product may directly threaten life or injury. Without limitation to the foregoing, WDT shall not be liable for any loss, injury or damage caused by use of its products in any of the following applications:

- Special applications such as military related equipment, nuclear reactor control, and aerospace
- Control devices for automotive vehicles, train, ship and traffic equipment
- Safety system for disaster prevention and crime prevention
- Medical-related equipment including medical measurement device

Accordingly, in any use of SanDisk-branded products in life support systems or other applications where failure could cause damage, injury or loss of life, the products should only be incorporated in systems designed with appropriate redundancy, fault tolerant or back-up features. Per the Terms and Conditions of Sale found in your order acknowledgement, the user of SanDisk-branded products in life support or other such applications assumes all risk of such use and agrees to indemnify, defend and hold harmless WDT and its affiliates against all damages.

Security safeguards, by their nature, are capable of circumvention. WDT cannot, and does not, guarantee that data will not be accessed by unauthorized persons, and WDT disclaims any warranties to that effect to the fullest extent permitted by law.

This document and related material is for information use only and is subject to change without prior notice. WDT assumes no responsibility for any errors that may appear in this document or related material, nor for any damages or claims resulting from the furnishing, performance or use of this document or related material. Absent a written agreement signed by WDT or its authorized representative to the contrary, WDT 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 EXPRESSLY WAIVES ALL GUARANTIES AND WARRANTIES OF ANY KIND ASSOCIATED WITH THIS DOCUMENT AND/OR RELATED MATERIALS, WHETHER EXPRESS OR IMPLIED, INCLUDING WITHOUT LIMITATION, ANY IMPLIED WARRANTY OF MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE OR INFRINGEMENT, TOGETHER WITH ANY LIABILITY OF WDT AND ITS AFFILIATES UNDER ANY CONTRACT, NEGLIGENCE, STRICT LIABILITY OR OTHER LEGAL OR EQUITABLE THEORY FOR LOSS OF USE, REVENUE, OR PROFIT OR OTHER INCIDENTAL, PUNITIVE, INDIRECT, SPECIAL OR CONSEQUENTIAL DAMAGES, INCLUDING WITHOUT LIMITATION PHYSICAL INJURY OR DEATH, PROPERTY DAMAGE, LOST DATA, OR COSTS OF PROCUREMENT OF SUBSTITUTE GOODS, TECHNOLOGY OR SERVICES.

No part of this document may be reproduced, transmitted, transcribed, stored in a retrievable manner or translated into any language or computer language, in any form or by any means, electronic, mechanical, magnetic, optical, chemical, manual or otherwise, without the prior written consent of an officer of WDT. All parts of the WDT documentation are protected by copyright law and all rights are reserved.

SanDisk and the SanDisk logo are registered trademarks or trademarks of Western Digital Corporation or its affiliates in the U.S. and/or other countries. microSD, microSDHC, microSDXC and SD marks and logos are trademarks of SD-3C, LLC. Other brand names mentioned herein are for identification purposes only and may be the trademarks of their respective holder(s). Performance will vary depending on your hardware and software components and configurations. Product specifications subject to change without notice. Pictures shown may vary from actual products. Not all products are available in all regions of the world.
The SanDisk Advantage

- More than 27 years expertise in NAND flash development and system design
- Vertically Integrated
- World Class technical and design support
- BOM control with PCN support

Contact Information
For all inquiries, please email:
oemproducts@sandisk.com
For more information, please visit:
www.sandisk.com