eMMC Write Amplification

Author: Mohammad Nada
Associated Part Family:S40FCxxx

1. Introduction

The NAND Flash is a popular storage media that dominates the non-volatile Flash market. NAND Flash's storage capacity, power efficiency, cost effectiveness, and scalable design make it an ideal choice for a wide range of applications. Flash densities have roughly doubled every year and manufacturers are pushing flash into smaller geometries and newer technologies. As a result, NAND flash is able to store more bits per cell to achieve higher densities and decrease the cost per gigabyte. eMMC Managed NAND provides simpler solutions and speeds time-to-market. The eMMC controller is embedded within the package and handles all NAND management operations such as wear leveling, bad blockmanagement, and ECC.

NAND Flash array contains one or more planes. Each plane is composed of a number of blocks, which are the smallest unit that can be erased. Each block is composed of several pages, which are the smallest unit that can be programmed.

The program operation is page based, while the erase operation is block based. When utilizing NAND flash, a block needs to be erased before it can be programmed again, all the pages in a block will be erased in an erase operation. If a piece of information needs to be updated in a block, the eMMC must completely erase the original data and move the updated information to a new empty block. Updating information in a block requires the eMMC to erase the original data and move the updated data to new block adding an extra program/erase (P/E) cycle. The added P/E cycles will have a direct impact on the device lifespan as it increases what is known as the Write Amplification (WAF).

2. NAND Flash Endurance

Historically, NAND Flash memory endurance has been calculated by the number of program/erase (P/E) cycles attributed to the type of NAND Flash memory. The number of program/erase (P/E) cycles of the NAND flash represents its lifespan. Every cycle degrades the memory cells by weakening the oxide layer of the cell when electrons are injected or extracted from the floating gate using what is known as Fowler–Nordheim (FN) tunneling. FN tunneling occurs when a high voltage is applied across the control gate or the substrate. A program operation involves injecting electrons into the floating gate to write a “0” while an erase operation involves extracting electrons from the floating gate which becomes a “1”. eMMC managed NAND can handle anywhere from several thousand to tens of thousands of P/E cycles. An eMMC maximum P/E cycle count can differ depending on the type of flash memory and the eMMC design. Newer technology like quad-level-cell (QLC) NAND offers higher bit density (four bits per cell, as opposed to SLC’s single bit per cell), but the P/E cycle count of these drives is much lower than single-level-cell (SLC) drives. That’s because programming procedures become exponentially more difficult with high-bit-density designs.

The challenge for increased MLC/TLC/QLC adoption is improving write endurance; there are several techniques vendors are currently using to meet the challenge.

Figure 1 demonstrates the number of P/E cycles for the different NAND flash technologies

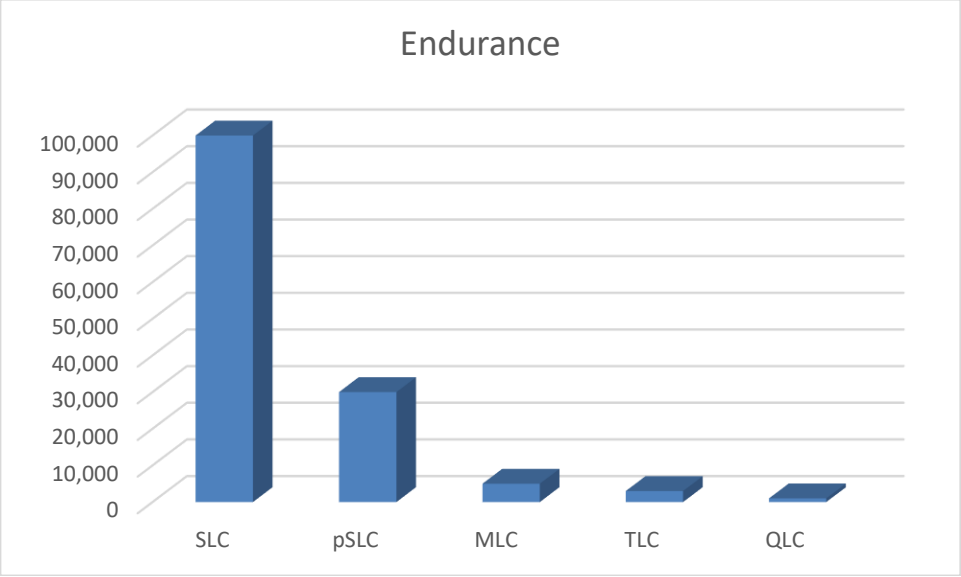

Figure 1: NAND Technology vs P/E Cycles (Typical)

3. NAND-flash Data writes

Data written to a NAND-based flash device goes through multiple layers. The data typically originates from the host, then moves through the file system to a buffer before it is finally written to the flash media.

At this level, there are two important concepts – a write page and an erase block. Each erase block is made up of several write pages. The smallest amount of data that can be written to NAND-based media is a write page. When application writes a data file – even if it is smaller than the write page size – an entire page must be written. Therefore, if the data file happens to be smaller than the write page capacity, the remaining area on the page is filled with bits of undefined data. Furthermore, if an application needs to expand the original file, a new NAND write page will be written with the expanded data. The process of consolidating data that has already been written to the flash is known as garbage collection.

It is common for managed NAND to have more data written to the flash memory than was originally sent by by the host system. This disparity is known as write amplification (WAF).

4. Write Amplification

Write amplification (WAF) refers to the phenomenon that occurs in the eMMCs in which the amount of data written to the memory device is greater than the amount of information being stored. Measuring WAF is usually done in ratio format by comparing the number of writes that are committed to the flash memory to the number of writes that are coming from the host system. WAF is measured as the ratio of the amount of data stored versus the amount of data received from the host system.

For example, if 2GB of data was sent from the host system while 4GB was written to the NAND flash, the write amplification would be a factor of 2.

Since flash has a limited lifespan based on the number of P/E cycles to the NAND flash memory, write amplification is major factor to be considered for the eMMC. High WAF factor lowers the eMMC Endurance and impacts the write performance since the eMMC will be performing multiple writes or erases internal to the flash.

The WAF impact does not become an issue until the entire eMMC has been fully written once since space is available and there is no need to do any garbage collection at that point.

However, once every block on the eMMC is used, garbage collection and possibly TRIM commands are necessary to make space for new data. The eMMC will then experience the highest levels of write amplification. The extent of the write amplification factor impact depends on few factors, such as, the size of the drive, the over-provisioning available space, and how the device is being used.

5. Factors Affecting Write Amplification (WAF)

- **Sequential writes vs Random writes:**

Different write processes have a different impact on drives and write amplification. Sequential writes aligned to page boundaries are the most effective form of writes to reduce the WAF. These writes occur when large amounts data are written at one time, in a sequential order. In this case, the data can fill up blocks with related sequential data. Therefore, when the system decides to erase the data, the entire block can be erased without write amplification. Writes with small amounts of data will lead to deleting and re-writing the existing data to create a new space for new data when the need arises, that will increase the write amplification.

However, with random writes, a process of read-modify-erase-write will occur for every new write cycle. Since the data is not stored on blocks sequentially, and when the need arises to create a free space for new data, blocks must be deleted and the existing data must be re-written to entertain the new data. This process will have a major impact on the WAF factor.

- **Free space Lowers Write Amplification:**

The NAND flash memory in solid state drives must be erased before it can store new data. Therefore, the eMMC's use a process called garbage collection (GC) to reclaim the space taken by previously stored data.

Having free space on the eMMC will keep the WAF lower and there are two methods to expand free space and lower the WAF:

1) Over- provisioning: refers to the creation of storage capacity that is not available to the host system and not included in the stated storage capacity of the eMMC.
   Such extra storage will be used for flash management which is used to store the extra copies of data during the garbage collection
2) Free space: more free space translates into lower write amplification.

When data is written randomly, the eventual replacement data will most likely come in randomly. Therefore, some pages of a block will be replaced (made invalid) and others will still be good (valid). During GC, valid data in blocks needs to be rewritten to new blocks. This produces another write to the flash for each valid page, causing higher write amplification. With sequential writes, generally all the data in the pages of the block become invalid at the same time. Therefore, there is no need to relocate data during GC and there is no impact on write amplification. If data from the host is compressed and de-duplicated before being sent to the eMMC, (such as PDF, ZIP or RAR files), the data writes to the eMMC will become more sequential. This will improve free space and lower WAF.
Additional free space enables write operations to be completed faster, which translates into a higher write speed and a longer eMMC lifespan.

Garbage Collection

Device level processes also increase write amplification factors and cannot always be controlled by the system designer. One example of such processes is garbage collection.
Data written to the eMMC is stored in pages, which are then grouped into blocks. As data is written to an eMMC, it may be written into different parts of the memory in different pages and different blocks, which complicates the process of erasing it.
To manage this, the process of garbage collection occurs where a flash block has a mix of valid and invalid data to be saved and invalid data to be erased.
In this process, the data that is to be saved is rewritten to another block, which has already been erased to makespace for new data. The block that is left with the invalid data to be deleted can be erased.
The Garbage collection process amplifies the number of program and erase cycles, and therefore has a big impact on the write amplification.

Wear leveling

Wear leveling, static and dynamic, can also increase the write amplification, especially for static data such as system files which are written once in a system. When the conditions are met and wear leveling is invoked, thisdata will be relocated to another erase block. Wear leveling is important as the flash media wears more evenly,however, it will contribute to an increase in the write amplification factor.

TRIM

TRIM is another feature that can help to control write amplification. In the event that the drive includes invalid Logical Block Addresses (LBAs), The TRIM command allows the LBAs to be erased when a file is deleted. Therefore, the systemwill no longer need to move and copy the LBAs during the garbage collection. The TRIM command helps to reduce the write amplification and increases the amount of free space available to the user.

6. Expected life calculations based on the write Amplification factor

Item	Value
Device Capacity	16GB
Write Endurance	3K program/erase cycles
Data written per day to device	2GB
Expected life w/ WAF=1 (16x3000)/(2x1)	24,000 days (1)
Expected life w/ WAF=5 (16x3000)/(2x5)	4,800 days (2)

Note (1): 16*3000/(2*1)
Note (2): 16*3000/(2*5)

7. Conclusion

Ensuring optimal system and data integrity remains a crucial factor in preventing device failures in the field. A factor like write amplification can cause issues with eMMC managed NAND, potentially decreasing the lifetime of the device.

While some of the factors that cause a higher WAF are out of the user's control, there are many techniques such as those listed in this application note that can reduce the WAF and prolong the life of the device.

Effective long-lasting data storage requires planning. Customized flash memory testing and analysis can provideclarity of the behavior of the flash device over the lifetime of the system. This knowledge can help embedded device manufacturers understand managed NAND capabilities and functions to prevent design issues and potential field failures.

8. Document History Page

Document Title: AN200009 - e.MMC Write Amplification Document Number: 002-00009				
Rev.	**ECN No.**	**Orig. of Change**	**Submission Date**	**Description of Change**
**	–	–	08/07/2024	Initial version