USING MLC FLASH TO REDUCE SYSTEM COST IN INDUSTRIAL APPLICATIONS

Chris Budd
SMART High Reliability Solutions
Gilbert, AZ 85233
Chris.Budd@smartH.com

ABSTRACT

Storage devices based on Multi-Level Cell (MLC) NAND flash can be found in almost all computer systems except rugged, industrial systems; even though MLC is less expensive and more dense than devices based on standard Single-Level Cell (SLC) NAND flash, MLC's lower write endurance and lower retention has led system designers to avoid using it. This avoidance is unnecessary in many applications which will never come close to the endurance limits. Furthermore, new processes are leading to storage devices with higher write endurance. System designers should review the specific use-model for their systems and can select MLC-based storage devices when warranted. The result is lower system costs without worry of data loss due to write endurance.

KEYWORDS

SSD, SLC, MLC, endurance, retention, temperature, file system, ruggedized

INTRODUCTION

System designers make choices with various system components and usually the decision comes down to cost or quality: does the system absolutely need the quality afforded by the higher priced component? Since system designers are not necessarily experts with the details of each component, they can repeat the success of previous systems by choosing the same or similar components. This process works while subsequent generations of the components are relatively similar. However, when a technological shift occurs, system designers need to revisit the available products that satisfy a system requirement. One such requirement is storage. Almost all systems need storage, and designers of rugged, industrial systems generally choose solid state storage devices such as Solid-State Drives (SSDs) with NAND flash. Previously, only Single-Level Cell (SLC) NAND flash was compatible with the requirements of rugged systems. Now recent improvements in the NAND flash technology and the NAND flash controller have allowed Multi-Level Cell (MLC) NAND flash to work reliably throughout the industrial temperature range (typically -40 °C to +85 °C). While MLC has a lower cost than SLC, and MLC has a higher density than SLC, system designers are rightfully concerned about the reduced write endurance and data retention of MLC versus SLC. System designers may understand retention, but they may not fully understand the endurance requirements of their system. This paper will review the differences between MLC and SLC, the recent improvements with MLC, and provide guidance to better understand endurance requirements.

1. DIFFERENCE BETWEEN MLC AND SLC

Before making rash decisions, a system designer needs to understand the fundamental differences between the choices along with their tradeoffs. In the case of rugged SSDs, the two main choices are SSDs built with SLC or SSDs built with MLC. The traditional preference, if not requirement, has always been SLC since MLC was not considered appropriate for rugged systems.

1.1 DIFFERENCES IN VOLTAGE LEVELS

While the cell structure of MLC and SLC is fundamentally the same, the main difference is in the interpretation of the voltage level of the cell. Single-Level Cell NAND flash consists of a single threshold level; if the voltage in the cell is below the threshold level, logic in the NAND flash interprets the cell to contain the bit value “1”; if the voltage is above the threshold, then the cell contains a “0” bit. In Figure 1, the single threshold level in the SLC NAND flash cell divides the voltage along horizontal axis into two distinct ranges each with an associated bit value, 1 or 0. Multi-Level Cell NAND flash consists of multiple threshold levels; MLC is generally identified with three threshold levels producing four possible ranges for the voltage of the cell; each range has a pair of bits associated with it as also shown in Figure 1. In one cell, MLC stores two bits while SLC stores only one bit; therefore, MLC has twice the density of SLC. While this increased storage density is desirable, it comes with the drawbacks described in the following sections.

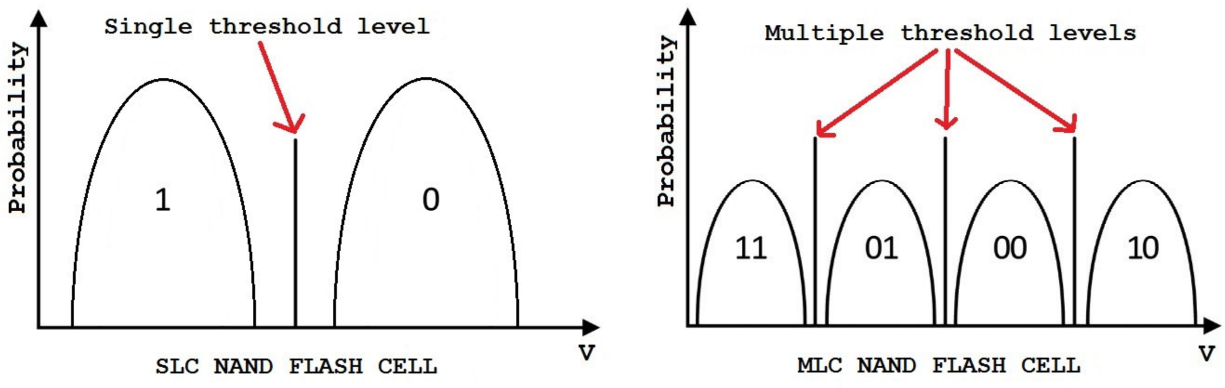

Figure 1: Threshold Levels in SLC and MLC

1.2 DIFFERENCES IN WRITE ENDURANCE AND RETENTION

The main drawback with MLC is its lower write endurance and reduced retention. The endurance of an SSD is generally defined to be the number of terabytes written (TBW) for a given capacity over the lifetime of the drive; retention is defined as the length of time the SSD can be powered off and still return the data without losing the data. A typical 240GB MLC drive has endurance of 100-400 TBW while a typical 240GB SLC has 20 times the endurance. Retention, on the other hand, varies depending upon drive age and temperature, but at 85 °C, an SLC drive can have well over 20 times the retention near the end of the MLC drive's endurance (73 days versus 2 days) [1]. Two difficulties with MLC cause the lower endurance and reduced retention.

As shown in Figure 1, there is essentially a 50% probability that an SLC NAND flash cell will have the voltage in the narrow range shown on horizontal axis below a particular curve. In an MLC NAND flash cell, the voltage range below each curve is still narrow, but the horizontal axis is more crowded with twice as many of these voltage ranges in the same cell structure. For any cells that end up near the thresholds, this crowding implies any minor variation of voltage in the cell could cause it to be interpreted as a different pair of bit values than was originally written. Since many of the modern NAND flash cells contain only a "handful of electrons" [2], it is easy to see how small disturbances could cause a small voltage variation and change the bits stored in the cell. In addition, after many program cycles, the curves tend to flatten out; with MLC, the curves cross not only the threshold levels, but the neighboring curves as well; therefore, as MLC wears out, it will have more bit errors without any external disturbance.

A second difficulty is that in order to write two bits to MLC, the flash controller must perform two program operations: once to set the least-significant bit, and the second to set the most-significant bit. This double program will greatly reduce the endurance of MLC compared to the single program required on SLC.

1.3 DIFFERENCES IN SENSITIVITY TO TEMPERATURE

All flash is subject to changes over temperature. For example, at hot temperatures, the endurance is greater since higher temperatures leads to fewer electrons getting trapped in the NAND flash cell, but consequently the retention is lower; the opposite is true for cold temperatures: reduced endurance, but longer retention [3]. Similarly, writing data to an SSD at one temperature extreme, and reading it back at a different temperature extreme causes problems for the SSD. Even though temperature affects all flash, it affects MLC more than SLC based on the reasons described previously.

2. RECENT IMPROVEMENTS

In spite of the difficulties with NAND flash, SSD users need not worry about them as the SSD and its NAND flash controller is designed to handle these difficulties and others. It is because of improvements to these NAND flash controllers, in addition to improvements to NAND flash technology itself, MLC drives exist at the industrial temperature range.

2.1 NAND FLASH TECHNOLOGY IMPROVEMENTS

When MLC was first introduced nearly 20 years ago, industrial temperature rated parts were not available. The goal of MLC was to increase the density of the NAND flash over SLC, and let SLC satisfy the other market segments such as those that need industrial temperatures. While NAND technology nodes get smaller and smaller, the densities of both MLC and SLC increase. New technologies like TLC and 3D NAND attract the attention and fabrication resources of the NAND flash vendors causing the vendors to be less interested in SLC and produce less SLC than they did previously [4]. Therefore, NAND flash vendors are now offering industrial temperature rated MLC parts.

2.2 NAND FLASH CONTROLLER IMPROVEMENTS

The early NAND flash controllers were not nearly as sophisticated as they are today. Improvements in the speed and density of these controllers have enabled them to accomplish more tasks in the same amount of time or shorter. Specifically, the NAND flash controllers have improved error correction algorithms and have added additional error correction bits, and the NAND flash has added additional space in which to store these extra bits; improved error correction helps to mitigate some of the problems discussed in previous sections including endurance and retention problems across the industrial temperature range. Some controllers include extraordinary error recovery techniques when the normal error correction algorithms fail. This additional and extraordinary error correction allows a controller to extend the endurance of the NAND flash. The controllers have also improved their wear-leveling algorithms so that they attempt to use all of the NAND flash blocks equally preventing a scenario where some blocks reach their endurance limit much sooner than other blocks. To help with retention, some controllers employ a read-scrubbing algorithm that periodically reads all the blocks and rewrites any block that is nearing the limit of the controller's error correction ability.

3. WRITE ENDURANCE REQUIREMENTS OF AN EMBEDDED APPLICATION

The improvements with MLC flash and NAND flash controllers are interesting, but a system designer still asks if an MLC SSD will have enough endurance for their application. The better question is, "How much endurance is needed?"

3.1 IMPACT OF A FILE SYSTEM

One important aspect of the endurance question is how the data is written. Will the system application write data through a file system, such as ext2 under a Linux operating system, or will it simply write data to the raw disk? A file system will increase the number of write operations to the SSD since the file system must manage its metadata like filenames and timestamps along with pointers to the files themselves. Some file systems are journaling, like ext3 and NTFS, meaning the file system writes a "journal" of its actions in addition to actually performing the action, which increases the number of write operations to the SSD even more. An alternative is to simply write to the raw disk without a file system; for example, if a system wants to record a stream of telemetry data, it can start at address 0 on the SSD and write to sequential addresses until the end of the steam or the end of the SSD. Not only does a raw write reduce the number of write operations to the SSD, it also helps the SSD manage its internal data mapping compared to the random nature of any file system; when using a file system, even if the application writes only one file sequentially, the file system must update its metadata and pointers causing the SSD controller to perform many additional internal read-modify-write operations to the NAND flash, and to perform frequent updates to its internal mapping tables.

3.2 TBW IN A REAL APPLICATION

The final answer to the endurance question depends on the system application. Below are three examples comparing the amount of data written against the endurance (TBW) of a typical MLC SSD.

- Map display: If the mapping information occupies most of a 240GB drive, the user could write a new map every day for 3 years before reaching 263 TBW, which would be near the endurance limit, 100-400 TBW, of a typical 240GB MLC SSD.

- Surveillance plane: If a surveillance plane records multiple streams of information to the SSD, the data access is more like random access than sequential access; because the SSD almost always has a slower random access data rate than its sequential access data rate, the maximum data rate of the SSD would be closer to 50 MB per second rather than 500 MB per second. During one 8-hour mission, where the application constantly writes 50 MB per second, the application would record close to 1.5 Terabytes (TB). Obviously, this application would require an SSD with a larger capacity, such as 1920GB, and one can expect the endurance to increase proportionally with the capacity producing a typical endurance of 800-3200 TBW. Therefore, one SSD could survive between 1.5 years to 5 years with one 8-hour mission per day.

- Ground-based transactional recorder such as in the base station of a cell phone tower: Although similar to the surveillance plane in recording small pieces of information randomly on the SSD, this recorder must be operational nonstop, 24 hours each day, 7 days each week; consequently, it will record 3 times more data per day than the plane did during its 8-hour mission. If this recorder has some mechanism to upload its data to an external server occasionally throughout the day, the drive could be written over and over at 50 MB per second – up to 4.5 TB per day. Using a large MLC drive, 1920GB, with an endurance rating of 3200 TBW, the replacement schedule for the drive should be just under 2 years.

Note that these examples assumed a worst-case amount of data written. A system designer would be wise to watch attributes from the drive, such as the attributes conveyed through Self-Monitoring, Analysis, and Reporting Technology (S.M.A.RT.), to determine a more accurate amount of data the application writes; many times the amount of data written is much less than the specified endurance of the SSD [5].

4. ADDITIONAL SELECTION CRITERIA FOR AN INDUSTRIAL APPLICATION

Endurance and retention are not the only requirements for an MLC SSD in a rugged environment such as those found with telemetry applications. These environments usually demand special features not found in commercial, consumer-class SSDs. Industrial temperature is an obvious difference, but other features could include a rugged enclosure capable of meeting MIL-STD-810G, special military erase sequences, write-protection, case isolation, interface connectors with 30 microinches of gold, etc.

4.1 DIFFERENT GRADES OF MLC

Another selection criterion for an industrial application could be the difference between MLC and enterprise-grade MLC (eMLC). The technology is the same between MLC and eMLC except that the eMLC flash is specially tuned to increase endurance. Even though greater endurance is welcome, it comes at the expense of retention. Typically in enterprise environments with uninterruptible power supplies, the application performs many small write operations leading to more emphasis on endurance than the length of time an SSD can retain data while powered off. Since most industrial applications cannot guarantee power, they require a reasonable balance between endurance and retention, and cannot use eMLC.

CONCLUSION

Historically, rugged, industrial systems required solid-state storage based on SLC. Even though SLC is more expensive than MLC, the reasons for this SLC exclusivity included temperature rating along with endurance and retention. However, recent improvements in NAND flash technology and NAND flash controllers allow MLC to exist in environments spanning the industrial temperature range. While retention is determined by the system designer, the endurance of MLC was the last obstacle to selecting MLC for these systems. Newer, larger capacity models allow for much more data to be written to the SSD than in years past, and most likely exceed the endurance requirements of many applications. This technological shift should encourage systems designers to review their specific application. They can not only compute the worst-case amount of data written, but also estimate the actual amount of data written using attributes from the drive. They may find the endurance they require is well below the endurance provided by SLC drives, and even below the endurance provided by MLC drives. Then the lower cost of MLC versus SLC allows the designer to save money on the overall system.

REFERENCES

[1] S. Phillips, "The Effects of Industrial Temperature on Embedded Storage," *RTC,* March 2015.

[2] P. Brace, "High-Performance and Large-Capacity Storage: A Winning Combination for Future Data Centers," in *Flash Memory Summit*, Santa Clara, CA, 2015.

[3] E. Tiomkin, "Industrial Temperature and NAND Flash in SSD Products," 24 July 2012. [Online]. Available: https://www.eeweb.com/blog/eli_tiomkin/industrial-temperature-and-nand-flash-in-ssd-products. [Accessed 25 May 2017].

[4] A. Lovinus, "SSD Storage: Is SLC NAND Extinct?," 28 September 2015. [Online]. Available: https://blog.neweggbusiness.com/components/ssd-storage-is-slc-nand-extinct/. [Accessed 26 May 2017].

[5] J. Handy, "Comparing DWPD to TBW," 1 June 2015. [Online]. Available: http://thessdguy.com/comparing-dwpd-to-tbw/. [Accessed 25 May 2017].

NOMENCLATURE

ATA or Advanced Technology Attachment: An interface standard from T13 (http://www.t13.org/) for storage devices typically using a parallel bus (PATA), or serial bus (SATA).

Flash: A non-volatile memory device using an array of transistors each with a floating gate to store a charge.

GB or Gigabyte: 10^9 bytes.

HDD or Hard-disk drive: Traditional mass storage device using a rotating, magnetic platter.

MB or Megabyte: 10^6 bytes.

MLC or Multi-Level Cell: A flash technology which stores more than one bit per transistor cell by using more than two voltage ranges on the floating gate; most commonly referred to two bits per cell using four distinct voltage ranges.

NAND: A high-density flash device usually with defect blocks marked by the factory; read and write operations must be done at a page level (several kilobytes), and erases must be done at an erase block level consisting of several hundred pages.

SATA or Serial ATA: A storage bus interface where the data is transferred serially rather than through parallel data wires as in previous ATA devices.

SED or Self-Encrypting Drive: A storage device that automatically encrypts all data written to the storage medium and decrypts all data read from the storage medium. No user or host intervention or support is necessary to manage the encryption key or perform the encryption.

SLC or Single-Level Cell: A flash technology which stores one bit per transistor cell by using two distinct voltage ranges on the floating gate.

S.M.A.R.T. or Self-Monitoring, Analysis, and Reporting Technology: A standard protocol for a host system to retrieve information about the drive status and health.

SSD or Solid-State Drive: A mass storage device typically using the same form factors as traditional hard-disk drives, but without the moving parts. An SSD typically stores data in SLC or MLC NAND flash.

TB or Terabyte: 10^{12} bytes

TBW or Terabytes Written: A measure of endurance of an SSD for a given capacity.