ADVANCED NAND MANAGEMENT TECHNOLOGY INCREASES DATA RELIABILITY AND EXTENDS THE LIFE OF SOLID-STATE DRIVES

White Paper

April 28, 2010
Version 1.0

This White Paper Describes the Advanced NAND Management Technology Integrated in the Fortasa Solid State Drive and Module Products

Please Contact Fortasa Sales for Any Additional Technical Information About Fortasa Products

4151 Middlefield Road
2nd Floor
Palo Alto, CA 94303 USA
888-367-8588
www.fortasa.com

Advanced NAND Management

Overview

One of the fastest growing semiconductor market segments is flash memory. Flash memory, and more specifically the NAND flash variant, offers unparalleled benefits of low-cost nonvolatile storage, small form factor, high density and low power consumption. In addition, the rapid rate of semiconductor innovation and technological progress is pushing memory costs lower and storage capacity higher. As the cost of NAND flash memory has rapidly decreased, high-volume consumer applications have embraced the benefits of flash, driving substantial production volume of the most cost-effective flash components. However, a major side effect of the dramatic cost reduction is also a reduction of NAND flash reliability. The parameters most often used to discuss flash component reliability are endurance (how many times the same cell can be programmed and erased) and data retention (how long the data can be retained over time).

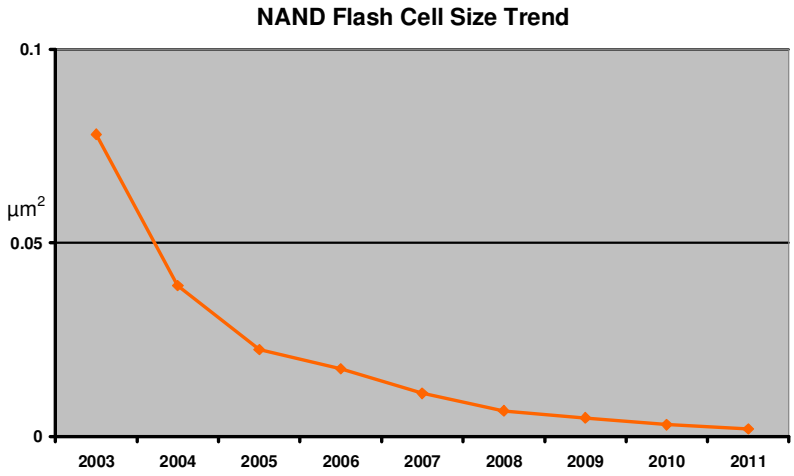
To compensate for the reduced reliability of the NAND flash components and to extend the useful life of the flash drive, multiple flash management techniques are designed into the NAND controller. Solid-state drive (SSD) products from Fortasa Memory Systems, Inc. contain patent-pending advanced NAND management technology including SMART-implemented unique features and advanced Write Filtering to increase data reliability and extend the useful life of the solid state storage product.

Introduction

Over the last few years the technological progress of NAND flash, the storage media in solid-state drives, has been unparalleled. In the tireless pursuit to lower the cost-per-bit of storage, NAND manufacturers have addressed the cost reduction challenge in two ways: first, by reducing the physical geometry of the manufacturing process (process shrink) and second, by utilizing advanced multi-level cell memory technology (doubling the storage capacity of a memory transistor).

NAND Flash Cell

Process shrinks reduce the physical size of a memory cell thereby packing more memory cells into the same physical space. Given that the physical size of the memory array can be directly translated into device cost, squeezing more memory cells into the same area increases the device density and correspondingly reduces the cost-per-bit of memory.



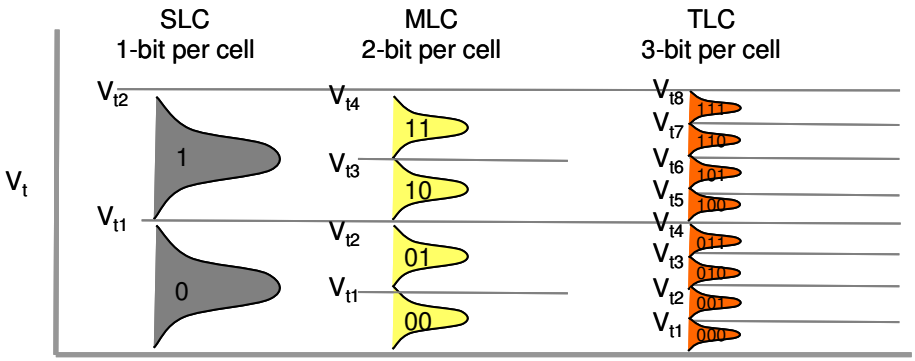
Picture 1: NAND flash cell size continues to reduce

A flash memory cell works on the principle that there is a distinguishable amount of stored charge inside the cell that corresponds to either a programmed or erased condition. Peripheral circuitry can check the status of the cell (read) and determine which state the cell is in. The greater the difference between the erased and programmed condition the easier it is to distinguish the cell state. As the process technology shrinks reduce the physical size of the memory cell, the amount of electrical charge that is possible to store inside diminishes, and differentiating between the programmed and erased cell condition gets an order of magnitude harder. In addition, the charge can leak out of the memory cell causing a false state to be read by the circuitry. This phenomenon is called data retention. As an intrinsic property of flash memory, data retention is also a function of how many program and erase cycles the memory cell has "endured". This is called flash memory endurance. Typically, as part of flash device qualification both data retention and device endurance are specified together.

As an example, JEDEC (JESD22-A117A, JESD47) specifies the relationship between endurance and data retention this way:

100% rated endurance	1-year data retention
10% rated endurance	10-year data retention

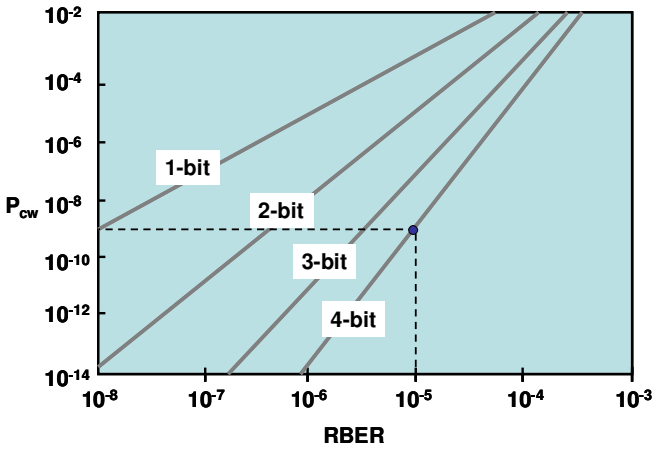
In parallel with shrinking process geometries, NAND suppliers have also incorporated a design technique that can store multiple bits of information in a single memory cell. Traditional single-level cell (SLC) memory stores a single bit of data in a memory cell. It sets a single voltage threshold to distinguish between a programmed or erased condition. Similarly, multi-level cell (MLC) technology stores 2-bits of information in a single memory cell and requires a distinction between four voltage threshold levels to get a proper reading. The next generation three-bit-per-cell (TLC) and quad-bit-per-cell (QLC) technologies will require distinguishing between eight and sixteen threshold voltage levels respectively. As the voltage levels are squeezed closer together it gets correspondingly tougher to identify the proper state of the memory cell to perform a correct reading of the stored data.



Picture 2: Voltage Levels for SLC, MLC and TLC flash technologies

The statistical likelihood of memory cell failure is stated as Raw Bit Error Rate (RBER) and is equal to the number of failed bits divided by total bits. RBER is statistically calculated by the NAND flash manufacturers for each specific NAND device.

The technique to improve the integrity of read data from the inherent RBER of a NAND device is called Error Correction Code (ECC). The ECC algorithm that is embedded in the flash memory controller calculates a mathematical codeword based on the programmed data. The codeword is then programmed in a separate section of a memory array. During a data read, the flash controller compares the read data to the codeword and mathematically corrects the failed bits, if necessary. The more robust the ECC, the more reliable is the data read from the NAND device.



Picture 3: Controller ECC improves bit error rate

As shown in Picture 3, for a given NAND device with a relatively high RBER of 10^{-5}, a controller with 4-bit ECC capability can improve the bit error rate to 10^{-9}. Solid State Storage products from Fortasa Memory Systems, Inc. contain up to 24-bit ECC, thus significantly improving the reliability over intrinsic RBER of NAND flash.

Flash manufacturers mandate the minimal requirement for the ECC correction depending on the semiconductor process and flash technology used in the NAND device as specified.

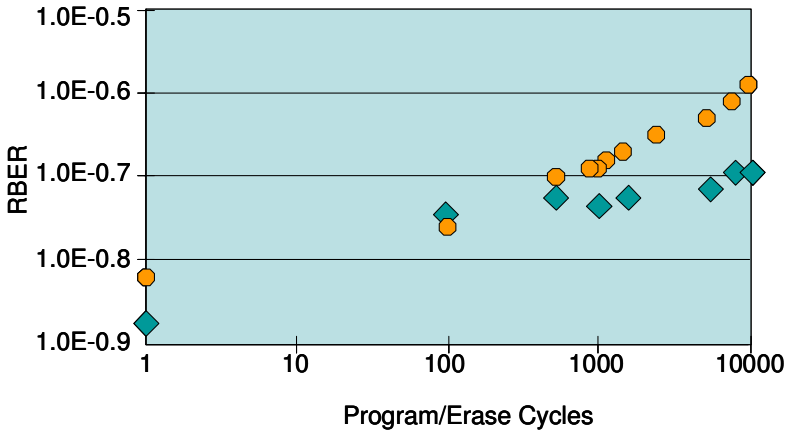
NAND Process Geometry		Rated Endurance w/1yr Data Retention
5xnm and above	SLC	100K cycles with 1-bit ECC
4xnm	SLC	100K cycles with 4-bit ECC
3xnm	SLC	100K cycles with 8-bit ECC
5xnm	MLC	**10K** cycles with **4-bit** ECC
4xnm	MLC	**10K** cycles with **8-bit** ECC
3xnm	MLC	**5K** cycles with **24-bit** ECC

Table 1: Smaller NAND process geometries yield lower data retention

Fortasa storage products use SLC and MLC NAND with an up to 12-bit per 512 Byte page ECC engine, offering significantly greater reliability than the flash manufacturer's requirement for embedded designs.

NAND Endurance

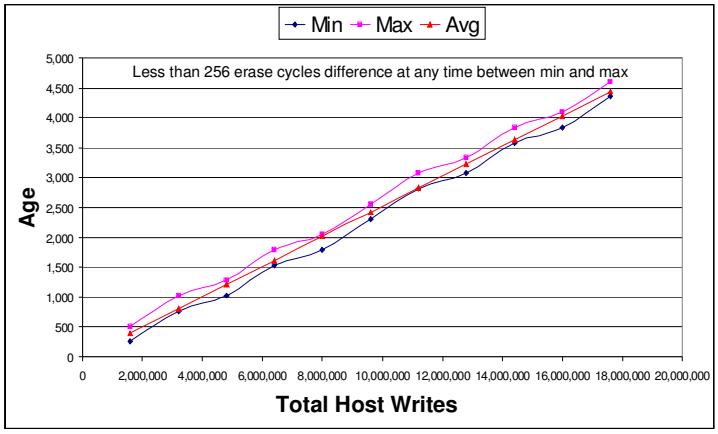
As mentioned before, NAND flash's data retention is directly proportional to the number of program and erase cycles exerted on the memory cell. Unlike hard disk drives (HDD), which do not degrade with continuous program/erase cycles, flash memory cells "wear out" with frequent use. Therefore, writing (programming and erasing) to the same physical address location continuously would overuse that memory location. The graph below shows how the RBER of NAND flash from two vendors increases as the number of program/erase cycles is increased.



Picture 4: Examples of increasing RBER with frequent program/erase cycles

To address the endurance limitation, the Fortasa SSDs utilize advanced data wear leveling techniques that spread the memory cell usage across the whole address space of the device. To enable wear leveling, Fortasa's NAND controller maps the address location requested by the host to a different physical location of the flash media. The controller keeps a mapping table that translates the host address location to a separate physical location in the flash. So potentially a continuous write command to the same logical location would actually be stored in any available address location on the flash media completely transparent to the host.

In addition, the NAND Controller's wear leveling algorithm keeps track of sector aging of all memory sectors and moves static resident data to a more utilized sector to spread the sector usage across the memory space. This proprietary algorithm allows a much more even wear of the flash media and provides more than 17 million host writes to the Fortasa Storage product before the endurance limitation becomes critical. The graph below shows that, during the 17 million host writes, the difference of minimum and maximum age (i.e. erase cycles) of blocks within an SSD is maintained at less than 256 erase cycles while the intrinsic endurance is 10K erase cycles. The graph also shows that after 17 million host writes, the SSD has consumed less than 5K, or 50%, of the intrinsic MLC Flash endurance cycles.



Picture 5: Fortasa SSD keeps the age differences among all blocks to be within 256 erase cycles

While the total number of physical program/erase cycles to the Fortasa SSD can surpass 17M, the actual number of writes from the host is less than that due to the phenomenon called Erase to Rewrite Factor.

Write Amplification Factor Causes Premature Aging of the SSD

From the host standpoint, the operating system (OS) follows the HDD convention and writes data in multiples of 512 Bytes or one sector of data. However, the smallest unit of data that can be written to a flash media is a *page*, which varies depending on the type of flash component used from 2 KBytes (4 sectors) to as many as 8 KBytes (16 sectors) to date. Furthermore, flash memory must first be erased before it can be successfully written, and the smallest erase unit of memory is a *block*, which consists of 128 or more pages (512+ sectors). The incongruity between a minimal write sector and an erase block multiplies the number of host write cycles, as every time a used block needs to be erased at least 128 pages of data may need to be written to a different flash location. Similarly, every time a single sector of 512 Bytes is updated by the host, a full programming page may need to be programmed into flash.

An OS write consists of updating not just the data sector but also the FAT (File Allocation Table) links to the data, thus amplifying the amount of program/erase cycles to the flash. Given the fragmented nature of FAT entries and potential data updates, the amplification factor of inefficient utilization of program/erase cycles to the flash would substantially reduce the inherent flash drive endurance rating and subsequently drive reliability.

Fortasa's NAND management technology contains a Write Filter that reduces the number of repetitive writes that can cause reduced drive endurance. The Write Filter accumulates repetitive writes to the same logical locations in order to filter out unnecessary block erase operations. A simple single-thread filter is not sufficient to reduce the drive wear as a typical write operation involves updates to both data and FAT entries. Utilizing a multi-thread Write Filter, Fortasa's NAND controller assigns multiple filters matching multiple erase blocks preventing extraneous cycling of FAT and data entries.

SSD Life Estimate Calculation in TBW (TeraBytes Written)

A critical part of the selection of the proper flash storage system is the ability to predict the expected minimum life of the product against an application-specific usage model. If the reliability of the storage system is rated higher than the host system requirement, then the selected component would function reliably through the useful life of the system. However, it is critical to be able to understand the usage model of the host system and compare it to the memory system reliability specification under the same usage.

As an example, the Fortasa SSD's useful life can be estimated under the following mathematical model:

1. **Wear leveling efficiency (WLE)** is defined as the ratio of the average number of erases on all blocks to the maximum erases on any block.

2. **Write Amplification Factor (WAF)** is defined as the average number of program and erase cycles for the SSD blocks divided by the number of drive rewrites, where drive rewrite is considered the host data written divided by the SSD capacity. WAF will depend on the operational algorithms of the NAND controller, the page size of the NAND device and the workload, and it may vary over the lifetime of the SSD.

3. **SSD capacity** is defined as the total specified capacity of the chosen SSD.

4. **NAND Endurance** is defined as the minimum endurance rating as specified by the NAND flash vendor (typically 100K cycles for SLC and 10K cycles for MLC NAND).

Under the above assumptions, the Fortasa SSD life estimate can be calculated in **Host-Write in TeraBytes Written (TBW)** by the following equation:

Host-Write (TBW) = [(NAND Endurance) x (SSD Capacity) x (WLE)] / WAF

Consider three usage models for typical OEM application:

Example 1: High Definition (HD) Video Recording – Mostly Sequential Writes of Large Files

Typical usage can be defined as follows:

- Data Transfer Size: 256KB
- Video File Size: 1.8GB per recording in HD format
- Total Recording: 5,000 times
- Maximum Required TBW: 9 TBW

Assuming the following empirical parameters for SSD, considering specific firmware and NAND flash:

- WLE = 0.9
- WAF = 1.1
- SSD Capacity = 16GB
- Minimum NAND Endurance Cycle = 5K

Using the life estimate formula above, one can calculate the minimum SSD life estimate for this application to be:

Host-Write = [(5K) x (16GB) x (0.9)] / 1.1 = 65.5 TBW

For this customer example, utilizing the provided host usage model, the calculated minimum life of the 16GB SSD of 65.5 TBW substantially surpasses the maximum customer requirement of only 9 TBW. It can therefore be deduced that the SSD solution offers about 7x the maximum required life expectancy for the customer in this application.

Example 2: Computing Application – Both Sequential and Random Writes of Various Size Files

Typical usage can be defined as follows:

- Total File Size Transferred Daily: 20GB for typical file size
- Total Product Life: 3 years
- Maximum Required TBW: 22 TBW

Assuming the following empirical parameters for Fortasa SSD, considering specific firmware and NAND flash:

- WLE = 0.9
- WAF = 3
- SSD Capacity = 16GB
- Minimum NAND Endurance Cycle = 5K

Using the life estimate formula above, one can calculate the minimum SSD life estimate for this application to be:

Host-Write = [(5K) x (16GB) x (0.9)] / 3 = 24 TBW

For this customer example, utilizing the provided host usage model, the calculated minimum life of the 16GB SSD of 24 TBW surpasses the maximum customer requirement of 20 TBW.

Example 3: Point of Sale – Mostly Random Writes of Small Size Files

Typical usage can be defined as follows:

- Data Transfer Size: 4KB
- Total File Size Transferred Daily: 180MB, 15MB/hr @ 12hr/day
- Product Life: 5 years
- Maximum Required TBW: 0.33 TBW

Assuming the following empirical parameters for Fortasa SSD, considering specific firmware and NAND flash:

- WLE = 0.9
- WAF = 16
- SSD Capacity = 2GB
- Minimum NAND Endurance Cycle = 5K

Using the life estimate formula above, one can calculate the minimum SSD life estimate for this application to be:

Host-Write = [(5K) x (2GB) x (0.9)] / 16 = 0.56 TBW

For this customer example, utilizing the provided host usage model, the calculated minimum life of the 2GB SSD of 0.56 TBW surpasses the maximum customer requirement of only 0.33 TBW

The examples above show that the minimum life estimate of a SSD is dependent on the host system usage model, but with advanced wear leveling and write filtering technology, Fortasa Solid State Storage products offer substantially greater reliability than required by the most demanding applications.

Product Life Monitoring

Even with all the sophisticated technology to alleviate shortfalls of the intrinsic reliability of the NAND flash, under extensive usage, the SSD will fail at some point in time. Up to now, there was really no method to predict this point of failure. Most computer users are intimately familiar with the blue screen of Microsoft Windows that shows up when the HDD fails. The blue screen failure event is random and unpredictable. To counter this catastrophic failure, most IT departments evangelized frequent back-ups of critical data.

With much more sophisticated NAND flash controller technology, one can now monitor the operation of the Fortasa SSD and predict the potential point of failure before the end of the rated product life. This capability allows an OEM a planned back-up, a graceful system shutdown and the replacement of a potentially vulnerable drive.

The Self Monitoring SMART command implementation from SST allows monitoring of multiple critical parameters of the usage of SSD. These parameters include number of available spare blocks, and maximum and minimum age of blocks. A user can set the monitoring routine to check these parameters at certain frequency and provide feedback to the host if a critical condition has been reached. When the critical parameter approaches a pre-specified maximum, a planned system maintenance and drive replacement service call may be scheduled.

Conclusion

NAND flash has become the storage of choice in consumer, industrial, and computing applications, offering significant benefits of high performance, low cost, small physical size, and low power consumption. However, un-managed raw NAND flash has side effects of limited reliability and finite data retention. Advanced NAND management technology designed into Fortasa's NAND controller and derivative Solid State Storage products improves the reliability of the NAND-based solution and extends the rated life of SSD products.

Revision History

Revision	Date	Description	Comments
1.0	04/28/2010	Initial Release	

Copyright © 2010 Fortasa Memory Systems, Inc. All Rights Reserved.
Information in this document is subject to change without prior notice.
Fortasa and the Fortasa logo are trademarks or registered trademarks of
Fortasa Memory Systems, Inc.
Other brands, names, trademarks or registered trademarks may be claimed as the property of their respective owners.