M-Systems
Flash Disk Pioneers

White Paper

Implementing MLC NAND Flash for Cost-Effective, High-Capacity Memory

Written by: Raz Dan and Rochelle Singer

JANUARY 2003
91-SR-014-02-8L, REV 1.0

Introduction

Multi-Level Cell (MLC) technology greatly reduces flash die size to achieve a breakthrough cost structure. It does this by storing 2 bits of data per physical cell instead of the traditional 1 bit per cell, using Binary flash technology. But the increased density of the MLC flash media has grave consequences in terms of data reliability and performance. A number of flash vendors, with varying degrees of success, have made attempts to implement MLC technology on selected flash platforms while overcoming its limitations.

Recently, Toshiba introduced MLC NAND technology. Although NAND constitutes a particularly good building block for MLC technology due to its high erase and write performance and high density (small size), MLC NAND is practically unusable for local data and code storage due to the degradation in data reliability and performance. x2 technology, implemented in M-Systems' DiskOnChip G3, combines advanced hardware features and customized software algorithms to boost reliability and performance levels to rival and even exceed those of 1 bit per cell devices.

This paper discusses the MLC solutions available on the market today, their benefits and limitations, and the hardware and software innovations of x2 technology that overcome MLC NAND limitations. It concludes with a discussion of how the combination of MLC and x2 technologies is enabling M-Systems' Mobile DiskOnChip® G3 flash disk to function as the most cost-effective, fast and reliable memory solution in mobile handsets and connected devices available on the market today.

Background

As smartphones, Personal Digital Assistants (PDAs), Set-Top Boxes (STBs) and other connected devices offer users more and more functionality and personalization options, the storage requirements of these devices have become substantially greater. For example, 2.5G terminals now incorporate 128Mbit (16MByte) or even 256Mbit (32MByte) of flash memory, compared to 16 to 32Mbit (2 to 4MByte) in 2G terminals. Users will enjoy designs based on standard operating systems with PC-like functionality and operational look-and-feel, support for multiple software applications and more sophisticated hardware, such as color screens, and a greatly increased area to store a mix of audio, video and text files. Despite these increasing storage requirements, the demand for sleek packaging, particularly in the cell phone market where small size and low weight are critical design elements, must also be met.

In an attempt to gain grounds in these highly competitive markets, vendors of flash memory are trying to squeeze more and more capacity into constantly shrinking silicon dies, thereby optimizing both size and cost benefits. While the obvious way to achieve this is to reduce the manufacturing process size, few vendors have found a way to pack more information into a single memory cell. The most mature of such technologies is Multi-Level Cell (MLC). Four levels of voltage are stored in a single cell (thus two bits per cell), as opposed to the traditional Binary flash technology, which stores two voltage levels (one bit per cell). The main challenges facing this technology are to program and sense the correct voltage level accurately and quickly.

No single company has managed to live up to this challenge on its own, although a few companies have introduced products that implement MLC technology with varying levels of success: Intel with NOR flash, Hitachi with AND flash and most recently, Toshiba with NAND flash. MLC flash was first mass-produced by Intel in 1999 with StrataFlash. This product succeeds in doubling the capacity

of NOR flash and achieving barely adequate reliability, but it has serious limitations: its performance is far slower than standard NOR flash.

NAND flash appeared to be the ideal media for data storage, due to its high-speed erase and write, high density (thus high capacity) and small size, as compared with NOR and AND devices. Based on these promising characteristics, Toshiba chose NAND flash as the basis on which to implement MLC technology. Toshiba's first MLC NAND product, just introduced in December 2002, offers up to a 50 percent decrease in die size compared to standard NAND, and about a 70 percent decrease in size, compared with competing NOR MLC products.

However, NAND flash itself is not a perfect media. It contains a large number of randomly scattered bad blocks, requires on-the-fly error correction, and uses a non-standard I/O interface, making it difficult to integrate. These limitations are dramatically worsened in MLC NAND, along with a slower programming time (compared to standard NAND) and a different software interface. The combination of these characteristics makes MLC NAND all but unusable as a stand-alone local data storage solution.

M-Systems' x2 technology, selected by Toshiba to enable their MLC NAND technology, implements reliability, performance and media management enhancements to perfect MLC NAND - without the need for a full scale controller (e.g., ATA or SCSI). The combination of MLC NAND and x2 technology in Mobile DiskOnChip G3 brings smartphones, STBs and other embedded systems the most cost-effective flash disk.

Comparing Binary and MLC Flash Technologies

Basic Flash Technology

Figure 1 shows the basic structure of a flash memory cell, which is similar to a standard MOS transistor. However, unlike a standard transistor, a flash cell must be able to retain charge after power removal in order to permanently store data. To accomplish this, a layer called the floating gate is added between the substrate and the select gate. The floating gate is isolated from the substrate and the select gate by layers of oxide.

A transistor can be biased (voltage can be applied to the source, drain, gate and substrate) to optionally conduct a current between its source and drain. The voltage level at which the transistor conducts is called its threshold voltage (V_{Th}). The transistor conducts only if the voltage between the select gate and source (V_{GS}) is larger than V_{Th}. Adding/Removing charge to/from the floating gate modifies the V_{Th}. To determine if the floating gate is charged, two conditions must be met: a specific V_{GS} must be applied to the cell and the circuit must be capable of sensing if the transistor is conducting. These are the basic elements needed to implement flash data storage.
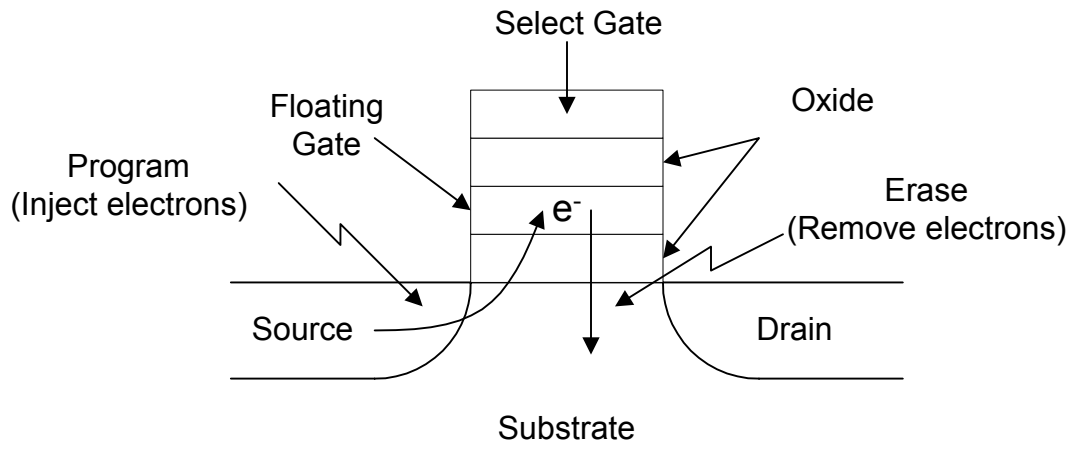
Figure 1: A Basic Flash Cell

Binary and MLC Technologies

In flash devices that implement Binary flash technology, there are two possible ranges for V_{Th}. MLC technology can have several valid ranges for V_{Th}, instead of just two. The first implementation of MLC uses four voltage levels (see Figure 2). Each state is mapped to one of four combinations of two bits. Therefore, the cell can store two bits of data.

Figure 2 also shows some of the complexity caused by the migration from Binary flash to MLC. The programming and erase processes become more complicated since the circuits must maintain tighter V_{Th} tolerances. This translates into longer program and erase times, and a more complicated read process.
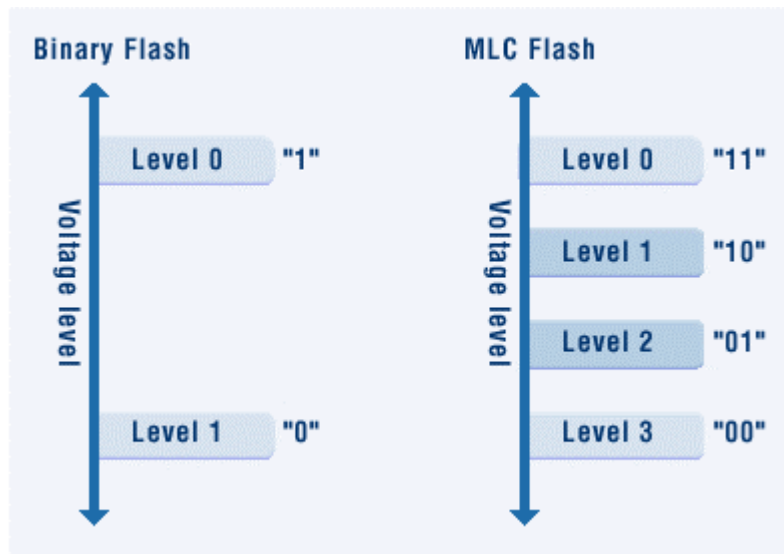

Figure 2: Voltage Level Comparison between Binary and Flash Technologies

MLC Benefits and Limitations

MLC high-density design innovations reduce the silicon die size, which is the major element contributing to overall device cost. For MLC NAND, this reduction in size and cost is greatest in capacities of 256Mbit (32MByte) and higher, where the die can be as small as 50 percent of the size required to provide the same capacity Binary flash device. The savings must be measured both in dollars and space, particularly for the cell phone market where every millimeter of board real estate can have an impact on the size of the end-user product and, ultimately, on market success.

But these very same high-density design innovations introduce three, major areas of design limitations as compared with Binary flash:

- Data reliability

- Performance

- Flash management

This section discusses these areas in order to lay the groundwork for understanding how x2 technology overcomes the associated problems.

Data Reliability

As shown in Figure 2, a Binary flash cell must distinguish between 2 voltage states, whereas an MLC flash cell must distinguish between 4. Since both Binary and MLC-based devices use a voltage window with a similar size, the distance between adjacent voltage levels in MLC is much smaller than in Binary flash. This reduced distance has an impact on data reliability. Detecting the voltage levels in an MLC flash cell is a more precise and complex task than in a Binary flash cell, subject to a higher probability of error that can affect data reliability in both the short and long term.

Assuming that the probability of all types of errors in Binary flash is on the order of 10^{-10}, the overall probability of MLC flash errors is two orders of magnitude worse.

Long-Term Data Errors

Flash memory cells must provide long-term data retention capabilities to function reliably as a non-volatile memory device. In order to do this, the long-term stability of voltage levels is critical. Leakage to/from the floating gate, which tends to slowly change the cell's voltage level from its initial level to a different level after cell programming or erasing, may change the voltage level. This new level may incorrectly be interpreted as a different logical value. Due to the smaller distance between MLC levels than Binary flash levels, MLC flash cells are more likely to be affected by leakage effects and, consequently, more potentially prone to errors.

Program Disturb Errors

The program disturb effect, also called over program effect, causes a programming operation on one page to induce a change in bit value on another, unrelated page. In Binary flash technology based on a 0.16µ manufacturing process, the typical program disturb error rate is on the order of 1 bit error per 10^{10} bits programmed. This compares with an error rate on the order of 1 bit error per 10^8 bits programmed with MLC flash technology.

Read Disturb Errors

The read disturb effect causes a page read operation to induce a permanent, bit value change in one of the read bits. In Binary flash technology based on a 0.16μ manufacturing process, the typical read disturb error rate is on the order of 1 bit error per 10^6 repetitive reads of the page containing the bit.

Although MLC cells are more prone to such errors, the effect in actual measurements is less severe than in program disturb errors. The measured rate is on the order of 1 bit error per approximately 10^5 repetitive reads of the page.

Performance

MLC technology requires more time than Binary flash technology for completing the basic flash operations of reading a page into the flash buffer, writing a flash buffer into a page, and erasing a flash unit. Especially for write operations, raw flash comparisons indicate that MLC performance is only 25 percent that of Binary flash. But many factors other than raw flash speed influence performance, including: host CPU bus timing issues, error detection and correction, software algorithms employed by the device driver, file system overhead, patterns of file access by the user, bus cycles and more.

In fact, from the user's point of view, raw read or write times are totally irrelevant. What the user "feels" is how long it takes from when, for example, a long sequence of write commands is issued to the file system, until the requests are completed. To get a "true" measure of these times, the measurements should be performed under scenarios that duplicate the real world as closely as possible. This implies first filling the disk to almost full capacity, and then performing the measurements, taking into account the hidden mechanisms of the software interfacing the flash to the user (file system, device driver, etc.).

Sustained Read

When comparing sustained read performance values in real-world scenarios for Binary Flash with MLC, the gap lessens considerably: MLC performance is 98 percent of Binary flash performance. Operations that both Binary flash and MLC require to support a sustained read operation – such as running the driver code and the file system code, and accumulating bus cycles to support address, command, error correction code and control information – account for closing the gap.

Sustained Write

A comparison of sustained write performance for both technologies in real-world scenarios must take into account an additional factor: making room for new data when no free space is available. This means adding to the calculation the time it takes to erase a flash unit and, depending on the time it takes to manage the flash (using M-Systems' TrueFFS®, for instance, adds 5 percent of the time required to write a unit), this time as well. For Binary flash, these calculations result in a sustained write performance rate of 250KBytes per second on a low MIPS platform, or 4 μsec per byte for a typical mix of files, as compared with 172KBytes per second for MLC. (Note that the number of sectors per unit for MLC is twice the corresponding number for Binary flash.) When these figures are translated into percentages, MLC sustained write performance is approximately 69 percent of Binary flash write performance.

Write performance greatly varies according to the user's access patterns, mainly the average file size. For large files the rate is much higher (up to approximately 600 KBytes per second); for very small files it is much lower. Here, unlike in read operations, the time that is required for file system handling is more significant than device driver time, especially when dealing with small files. Bus cycle time for writing is practically the same as for reading. All the remaining time is spent on software overhead.

Flash Management

Because of MLC's unique architecture, pages can only be written sequentially, whereas in Binary flash they can be written randomly within the erase block. MLC also makes partial page programming impossible, as opposed to Binary flash technology that enables it. This means that the existing translation layers used by TrueFFS to support Binary flash devices, NFTL and INFTL, are unusable, since they rely on random page access. Sequential write only and the lack of partial page programming impose limitations on MLC that affect reliability as well as performance.

Overcoming MLC Limitations

Because MLC technology can potentially bring the industry breakthrough cost and size benefits for local data and code storage, M-Systems chose to take on the challenge of perfecting it by providing solutions to overcome MLC reliability, performance and flash management limitations.

x2 technology, customized by M-Systems specifically to meet this challenge, is a combination of algorithms, performance-enhancing innovations and flash management capabilities. Developed in cooperation with Toshiba, x2 technology is integrated seamlessly into the different modules of M-Systems' Mobile DiskOnChip G3 architecture and fully compatible with its TrueFFS technology for flash management. x2 technology includes reliability and performance improvements integrated into TrueFFS, the thin controller and the flash media itself, as shown in Figure 3. x2 technology cleverly balances software and hardware to keep reliability and performance at their peek while maintaining MLC cost and size benefits.
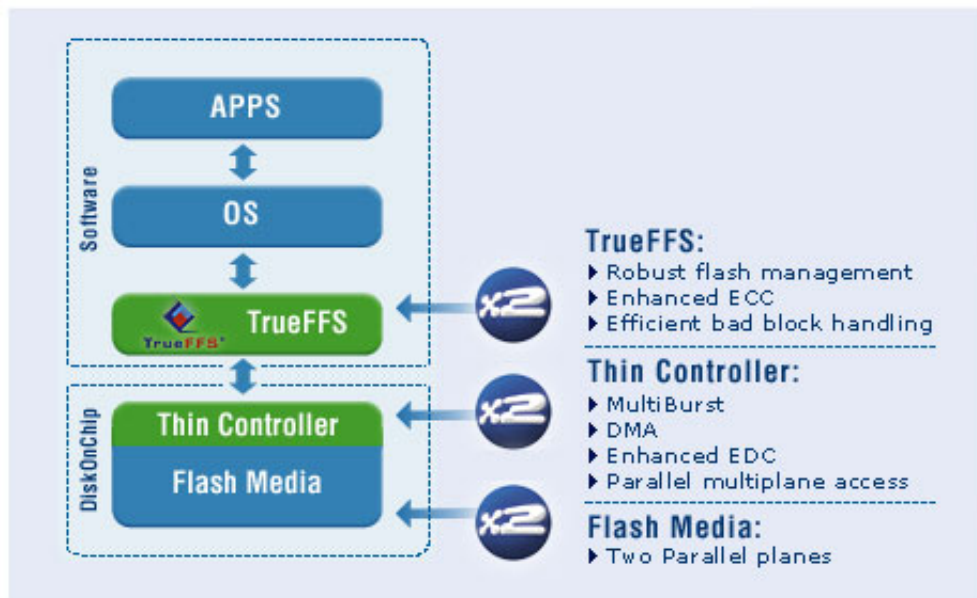

Figure 3: x2 Technology: Seamless Integration into M-Systems' Mobile DiskOnChip G3

Table 1 maps the various features of x2 technology against the three major areas of MLC limitations that they overcome. The remainder of this section explains how each feature achieves these enhancements in Mobile DiskOnChip G3.

Table 1: Overcoming MLC Limitations with x2-based Mobile DiskOnChip G3

x2 Technology Feature	Areas of MLC Enhancement		
	Reliability	Performance	Flash Management
TrueFFS			
Robust flash management	✔	✔	✔
Enhanced EDC	✔	✔	✔
Enhanced ECC	✔	✔	✔
Efficient bad block handling	✔		✔
Thin Controller			
MultiBurst		✔	
DMA support		✔	
Parallel multiplane access		✔	
Flash Media			
Two parallel planes		✔	✔

Robust Flash Management

To overcome MLC flash access and partial programming limitations that affect all three areas of MLC limitations, x2 technology uses a specially customized translation layer called Sequential Access Flash Translation Layer (SAFTL). SAFTL is incorporated seamlessly into M-Systems' TrueFFS. It maps each virtual unit into a chain of physical units, much in the same way that translation layers for Binary flash operate. However, unlike traditional translation layers, SAFTL does not implement one-to-one simple mapping between the virtual sector offset in the virtual unit and its physical location in the physical units. Instead, the data of a virtual sector can be in any location within the physical unit chain of its virtual unit. Each physical sector containing data also contains the offset of its corresponding virtual sector in its virtual unit.

SAFTL enables each physical unit to be filled sequentially, as required by MLC flash, starting from the first sector to the last. Each write request to the corresponding virtual unit is written to the next free physical sector, regardless of the virtual sector number requested to be written. When a physical unit is full and a new write request arrives, a new free physical unit is allocated and added to the chain. New unit allocation always occurs concurrently with writing a sector, so that sector data and unit control data can be written in one operation to improve performance.

Enhanced EDC and ECC

The Error Detection Code (EDC) and Error Correction Code (ECC) developed for x2 technology is based on M-Systems' highly effective combination used in previous generation DiskOnChip products. This system contains hardware-embedded EDC mechanism to detect errors on-the-fly and software-embedded ECC mechanism to reduce silicon size and cost. The combination of hardware and software results in the industry's most cost-effective data reliability for Binary flash. It corrects at least 2 errors per page without imposing performance penalties.

The EDC and ECC enhancements for MLC are capable of correcting up to 4 errors per page, using two industry-standard error codes: an extended Hamming code and a BCH (Bose, Chaudhuri and Hocquenghem) code.

The Hamming code can detect 2 errors per page and correct 1 error per page. The BCH code can detect 4 errors per page and correct an equal number, with a safety margin that enables it to detect 5 errors per page with a probability of 99.9 percent. This combination of codes provides an even higher rate of coverage than 2 bits per page provide for Binary flash technology.

It also ensures that the minimal amount of code required is used for detection and correction to deliver the required reliability without degrading performance. The entire thin controller occupies less than 5 percent of the die size for a 512Mbit device, of which only 15 percent is used for the EDC circuit to provide exceptional detection capabilities.

Efficient Bad Block Handling

x2 technology handles bad blocks, which can be randomly present in flash media, by enabling unaligned block access to two planes. Bad blocks are mapped individually on each plane, as shown in Figure 4. Good units can therefore be aligned or unaligned, minimizing the effects of bad blocks on the media. Without this capability, a bad block in one plane would cause a good block in the second plane to be tagged as a bad block, making it unusable. This customized method of bad block handling for two planes enhances data reliability without adversely affecting performance.
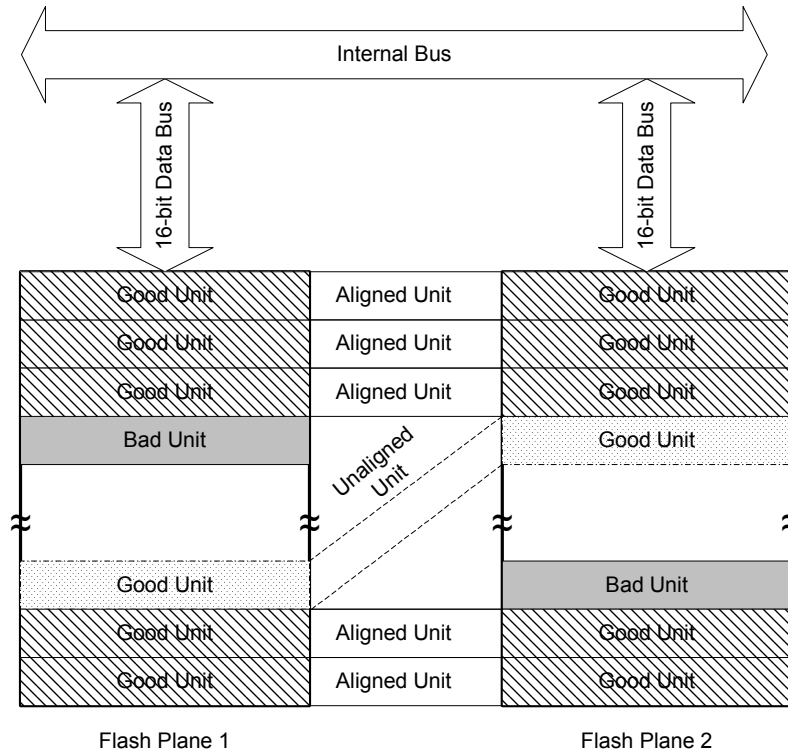
Figure 4: Unaligned Multiplane Bad Block Access

MultiBurst

To improve MLC read performance rates, x2 technology incorporates a feature called MultiBurst. MultiBurst enables parallel read access from two 16-bit planes to the flash controller, thereby achieving the desired output data rate for the host. The host accesses the first word of a page with a relatively slow access time, but each subsequent word with a very fast access time. Two cycles of 16 bits each are sent to the host at a clock rate set by the host rather than limited by flash operation, as shown in Figure 5.
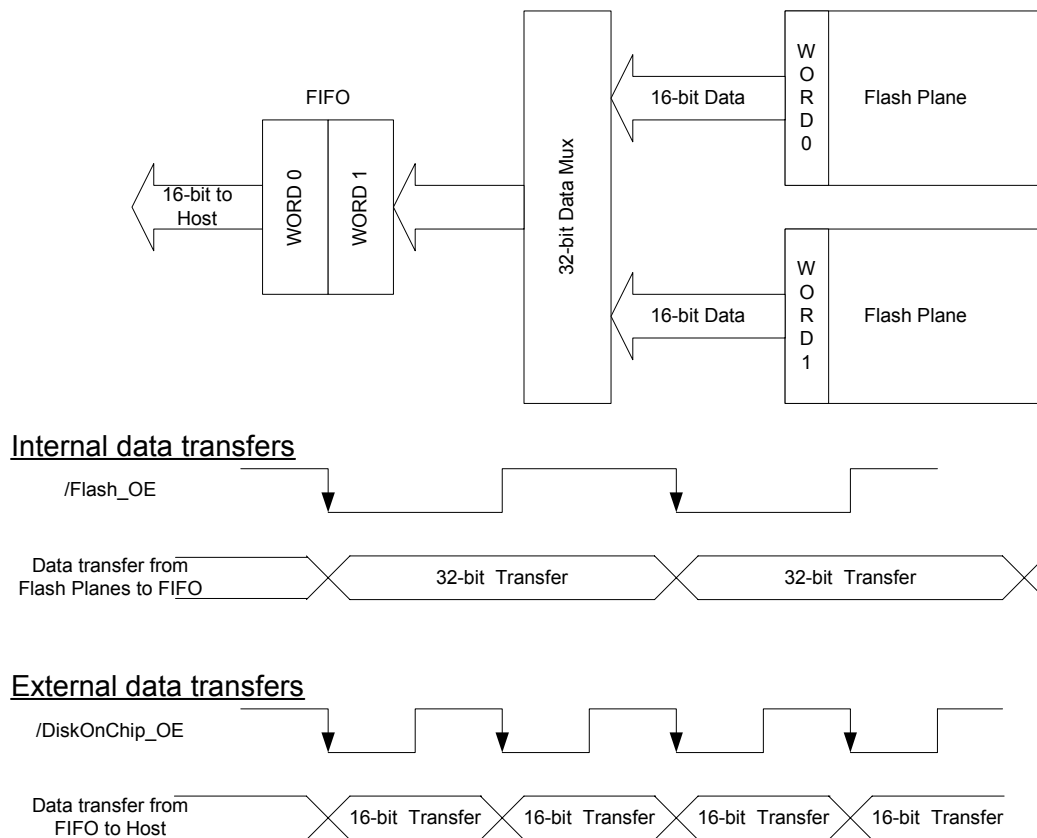
Figure 5: MultiBurst Operation

DMA Support

By enabling DMA operation, x2 technology reduces the CPU overhead. This is a particularly useful feature for transferring large files in support of Real-Time Operating Systems (RTOS). In addition, it can be used to enhance overall system performance by reducing boot time. In this case, the DMA mechanism is used to quickly transfer large blocks of code from the media into shadow RAM.

When comparing Mobile DiskOnChip G3 to raw flash products, such as Intel StrataFlash or AMD MirrorBit, this capability has at least a threefold benefit: increased performance, easier integration, and reduced external part count by allowing direct connection to a DMA controller without additional hardware.

Parallel Multiplane Access

As discussed earlier, the MLC flash media is built of two planes that can operate in parallel. This architecture is one of the most powerful, x2 technology innovations, doubling read, write and erase performance. Two pages on different planes can be concurrently read or written if they have the same offset within their respective blocks, even if the blocks are unaligned.

Power Consumption

M-Systems' Mobile DiskOnChip was designed for mobile systems that require very low power consumption. Therefore, the design incorporates power management features, such as Deep Power-Down mode, which consumes only 10 µA. Since the design is completely static (requiring no free-running clocks), it automatically goes into standby mode when not accessed. In addition, TrueFFS places Mobile DiskOnChip in Deep Power-Down mode at the end of every sector transfer. This design provides for a quick transition from Deep Power-Down mode to operational mode with minimal latency to minimize performance penalties.

Because x2 technology is seamlessly integrated into the existing DiskOnChip technology, power consumption levels for Mobile DiskOnChip G3 are equally as low. This is true despite the additional benefits of MLC and x2 technology.

Summary

The major improvements in flash NAND devices brought about by MLC technology are: much smaller size per bit, and consequently, a greatly reduced silicon size. These advantages come with added complexity in both device hardware architecture and device driver software. However, this document shows that x2 technology, by cleverly customizing the thin controller, TrueFFS and the flash media, provides a flash disk storage device based on MLC NAND that is as reliable and as fast as Binary flash devices in common use today.

Mobile DiskOnChip G3 512Mbit is M-Systems' first product to implement MLC NAND and x2 technology. This product meets OEM storage requirements for highly reliable, high performance, high-capacity data storage in 2.5G and 3G mobile devices, using a greatly reduced silicon size in a the industry's smallest BGA package, 7x10mm. Mobile DiskOnChip G3 is the most cost-effective memory solution available (Table 2). Future products that will use MLC NAND and x2 technology include Mobile DiskOnChip G3 in 256Mbit (32MByte) and 1Gbit (128MByte) capacities, as well as M-Systems' highly successful DiskOnKey keychain storage device.

Table 2: Comparing NAND Flash Alternatives

	Binary NAND	**MLC NAND**	**DiskOnChip G3 MLC NAND and x2**
Size	100%	~50% of Binary NAND	~53% of Binary NAND
Sustained write performance	~800KB/sec	~350KB/sec	~700KB/sec
Reliability	Occasional random errors	Frequent random errors	Perfect device

How to Contact Us

Website: http://www.m-sys.com

General Information: info@m-sys.com

Technical Information: techsupport@m-sys.com

USA
M-Systems Inc.
8371 Central Ave, Suite A
Newark CA 94560
Phone: +1-510-494-2090
Fax: +1-510-494-5545

Taiwan
M-Systems Asia Ltd.
Room B, 13 F, No. 133 Sec. 3
Min Sheng East Road
Taipei, Taiwan R.O.C.
Phone: +886-2-8770-6226
Fax: +886-2-8770-6295

Japan
M-Systems Japan Inc.
Asahi Seimei Gotanda Bldg., 3F
5-25-16 Higashi-Gotanda
Shinagawa-ku Tokyo, 141-0022
Phone: +81-3-5423-8101
Fax: +81-3-5423-8102

China
M-Systems China Ltd.
25A International Business Commercial Bldg.
Nanhu Rd., Lou Hu District
Shenzhen, China 518001
Phone: +86-755-2519-4732
Fax: +86-755-2519-4729

Europe and Israel
M-Systems Ltd.
7 Atir Yeda St.
Kfar Saba 44425, Israel
Phone: +972-9-764-5000
Fax: +972-3-548-8666