



Solid state drive technology for ProLiant servers

technology brief

Abstract.....	2
Introduction.....	2
Flash memory technology	2
Single-level and multi-level cell NAND flash.....	2
NAND architecture	3
Design of SSDs with flash memory	5
Wear leveling for increased SSD endurance	5
Over-provisioning NAND.....	6
HP solid state drives for ProLiant servers	6
Performance of HP server SSDs.....	6
Reliability of HP server SSDs.....	7
SSD futures	8
Summary	8
For more information.....	9
Call to action	9

Abstract

This technology brief is intended as a guide to emerging solid state storage technology, in particular, to the introduction of solid state drives designed for use in the ProLiant server environment. It provides a high-level overview of flash memory and the design methodologies involved in creating reliable solid state drives that use flash memory. Finally, it provides an analysis of both the performance characteristics and operating environment for solid state drives compared to those for the hard disk drives currently used in ProLiant servers. It is expected that the reader has basic knowledge of hard disk drives and their performance characteristics, as well as a familiarity with flash memory technology.

Introduction

Most users are familiar with flash memory from its use in consumer electronics products, including digital cameras, MP3 players, and USB flash drives. By combining flash memory with advanced controller technology, the industry has, over the past several years, been developing a new storage device based on flash memory—the solid state drive (SSD). Solid state drives interface with the host system using the same protocols as disk drives, but SSDs store and retrieve file data in flash memory arrays rather than on spinning media. Continuing advances in both flash memory and controller technology are, for the first time, enabling solid state drive designs that begin to meet the capacity, performance, and reliability requirements for use in server environments.

Flash memory technology

The majority of solid state drives are based on flash memory technology, a non-volatile computer memory that can be electrically erased and reprogrammed. Because flash memory lies at the heart of SSDs, it is important to understand the basic operating characteristics of this technology and how they influence SSD performance, reliability, and suitability for various application environments.

Flash memory is produced in two basic configurations, commonly referred to as NOR flash and NAND flash. Both NOR and NAND flash store information in arrays of floating-gate transistors referred to as “cells,” but they differ in how the cell arrays are organized and accessed. In NOR flash, cells are connected in parallel to the bit lines, allowing cells to be read and programmed individually. In NAND flash, cells are connected in series and therefore can only be read or programmed in series as a group.

The design choices made for NAND memory architecture mean that NAND memory arrays can be created with almost twice the density of comparable NOR memory, and at a lower cost. This has resulted in NAND flash becoming the predominate architecture in the marketplace, and it will probably remain so for the immediate future.

Single-level and multi-level cell NAND flash

There are two primary types of NAND flash technology: single-level cell (SLC) and multi-level cell (MLC). Single-level cell technology works by storing a single level of charge in each cell, representing a single bit of information. Multi-level cell technology stores one of four different charge states in a cell. This allows each cell to represent 2 bits of information, effectively doubling the memory storage density over SLC flash memory.

NAND flash memory using multi-level cell technology has quickly become the predominant flash technology used in the broader market for consumer products. However compared to SLC, MLC has the following characteristics that make it less desirable for creating the type of higher performance, high-reliability devices required for use in server storage (Table 1):

- Higher internal error rates caused by the smaller margins separating the cell states, necessitating larger ECC memories to correct them
- Significantly shorter lifespan in terms of maximum number of program/erase cycles
- Slower read performance and significantly slower write (program) performance

Table 1. Primary characteristics of SLC and MLC flash

	SLC flash	MLC flash
Random read	25 microseconds	60 microseconds
Serial access	50 nanoseconds	30 nanoseconds
Page program (write)	200 microseconds	800 microseconds
Maximum program/erase cycles	100,000 @ 1 bit ECC	5000 – 10,000 @ 4 bit ECC

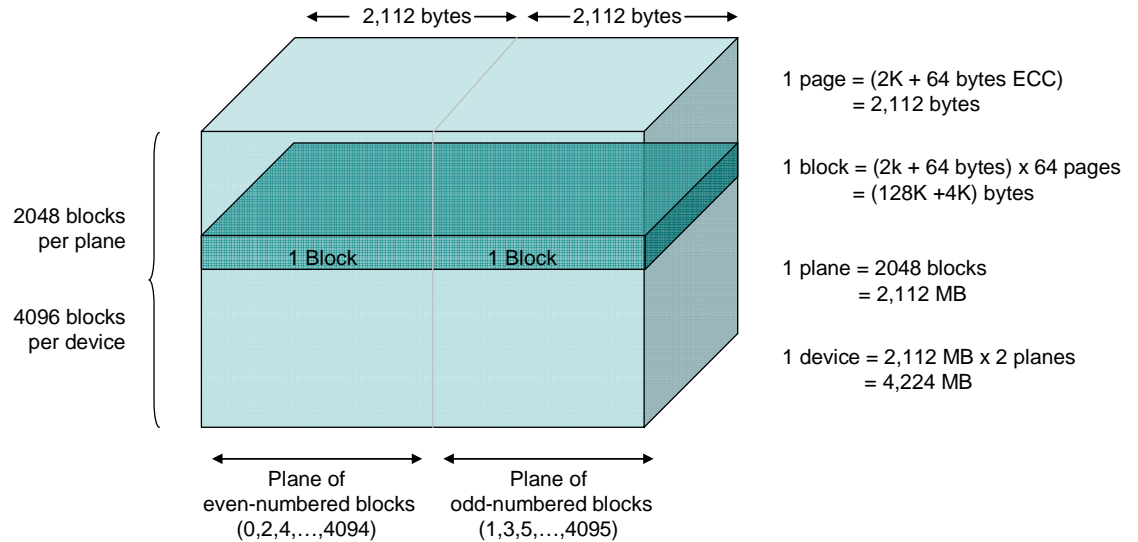
As Table 1 shows, MLC NAND flash has comparatively poor read and write performance. More importantly, SLC flash has a program/erase lifecycle—often referred to as endurance—that is ten to twenty times greater than that of MLC flash. MLC’s higher storage density will continue making it the predominant choice for use in lower cost and lower workload consumer devices. SLC NAND’s higher performance and better reliability are currently preferred to create solid state drives that can meet the requirements of server storage.

NAND architecture

NAND flash memory arrays are organized into pages and blocks. A page is the smallest organizational unit of a NAND array. Page size can vary between different NAND implementations but they are typically 2KB, 4KB or 8KB in size. Pages are then organized into blocks. Each block typically consists of 64 pages, although this too can vary between implementations. These will be referred to as NAND blocks throughout the remainder of this paper to differentiate them from the 512-byte logical block of the SATA/SAS interface.

SLC NAND can also be implemented in a two-plane architecture in which the device is divided into two physical planes consisting of the odd and even blocks respectively. Use of two-plane flash improves basic NAND performance by allowing two pages to be read or programmed concurrently. It also enables concurrent erasing of two blocks. Figure 1 shows the architecture for a 4-GB SLC NAND architecture consisting of 2K pages with 64 pages per block. NAND architecture continues to evolve at a fairly rapid pace, with 8K pages becoming common and four-plane designs on the horizon.

Figure 1. Basic organization of NAND memory



NAND flash has a very specific protocol for writing and retrieving information. NAND memory, unlike DRAM, must be accessed in discreet units. The smallest unit that can be read or written is the page. Unlike disk drives, pages that contain existing data cannot be directly overwritten with new data; they must first be erased. Adding to this complexity is the fact that NAND memory can only be erased in entire NAND blocks, which typically consist of either 64 or 128 pages. One of the more important tasks for any storage device built using NAND flash is effectively managing this asymmetry of the size of writes versus erases. Table 2 provides a list of these basic NAND operations and their execution times.

Table 2. SLC NAND flash operations

Operation	Execution time
Random page read	25 microseconds
Page program (write)	200 microseconds
Block erase	1500 microseconds

As Table 2 shows, writing to NAND flash is a considerably slower operation than reading from it. A page program operation is eight times slower than a random page read. A block erase operation, which is executed less frequently but is still part of the NAND programming overhead, is seven times slower than page program operation. Although many high level strategies are used to address this timing disparity, it is the primary reason that all NAND-based storage devices, whether USB drives or the more advanced solid state drives, have measurably better read performance than write performance.

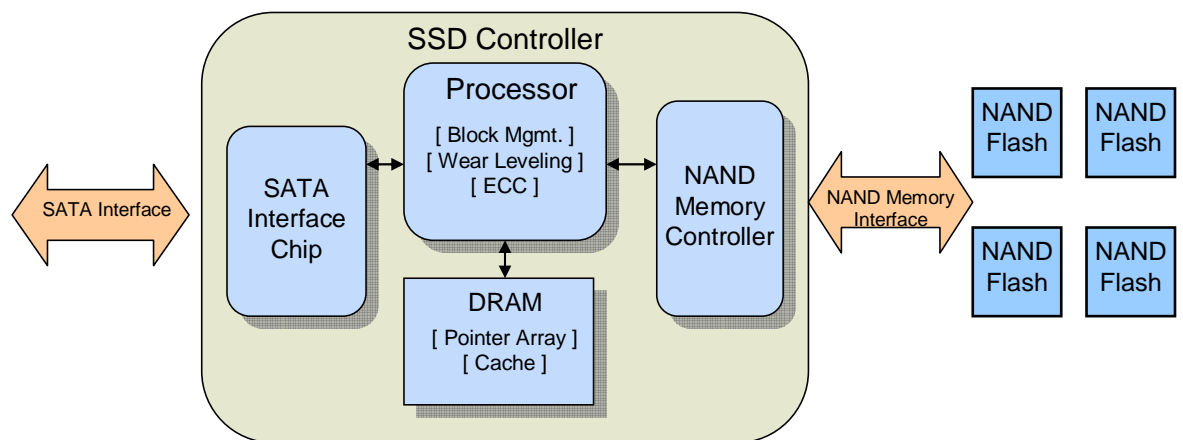
Design of SSDs with flash memory

Creating NAND-based solid state drives requires designing a drive controller subsystem that accomplishes the following tasks:

- Efficiently manages read and write operations to the NAND memory, including error handling and block management
- Enhances the native performance of NAND flash using management algorithms and RAM-based cache
- Maximizes the endurance, or lifespan, of the SSD by employing algorithms to minimize write/erase cycles to the NAND memory
- Provides the basic translation between the NAND read/write interface and the desired interconnect to the host, typically SAS or SATA

Figure 2 is a basic functional diagram for a typical SATA SSD. It includes the SSD controller section that provides all of the operational logic necessary to manage NAND flash memory and provide a standard SATA storage interface to the host server.

Figure 2. Functional diagram of a typical SATA solid state drive



Wear leveling for increased SSD endurance

Wear leveling is one of the basic techniques used to increase the overall endurance of NAND-based SSDs. Since NAND-based SLC flash supports only 100,000 lifetime write/erase cycles, it is important that no physical NAND block in the memory array be erased and re-written more than is necessary. However, oftentimes certain logical SCSI blocks of a SAS/SATA device may need to be updated, or re-written, on a very frequent basis. Wear leveling resolves this issue by continuously re-mapping logical SCSI blocks to different physical pages in the NAND array. Wear leveling ensures that erasures and rewrites remain evenly distributed across the medium, which maximizes the endurance of the SSD. To maximize SSD performance, this logical-to-physical map is maintained as a pointer array in high speed DRAM on the SSD controller. It is also maintained algorithmically in metadata regions in the NAND flash array itself. This ensures that the map can be re-built in the case of an unexpected power loss.

Over-provisioning NAND

The overall endurance and performance of an SSD can also be increased by over-provisioning the amount of NAND capacity on the device. On higher end SSDs, NAND can be over-provisioned by as much as 25 percent above the stated storage capacity. Over-provisioning increases the endurance of an SSD by distributing the total number of writes and erases across a larger population of NAND blocks and pages over time. Over-provisioning can also increase SSD performance by giving the SSD controller additional buffer space for managing page writes and NAND block erases.

HP solid state drives for ProLiant servers

First-generation HP server SSDs were introduced in late 2008 for use in specific BladeSystem environments and are available in 32- and 64-gigabyte capacities. HP server SSDs are designed to meet the higher standards required of storage devices for the server environment. At the same time, they provide the unique performance and reliability characteristics associated with SSDs in general.

Performance of HP server SSDs

HP server SSDs are interface-compatible with traditional disk drives connected to a SATA controller. This allows benchmarking and direct comparison of their external performance with that of disk drives to determine their suitability in various application environments.

The overall performance of a traditional disk drive is influenced by the disk access time, or latency, which is the total time required for the system to retrieve data from the drive. Disk drive latency is the sum of the seek time, rotational delay, and transfer time.

With SSDs there is no seek time or rotational delay. Latency is primarily a function of the memory access and transfer times combined with controller overhead. Given this fact and the knowledge of how NAND flash operates, we should expect the following to be true:

- Read operations in general should be faster on SSDs than write operations because of the relative slowness of NAND program (write) operations.
- Random reads on SSDs should be exceptionally fast compared to random reads on disk drives, since SSDs eliminate the seek time and rotational delay for each read operation.

Table 3 is a side-by-side comparison of the performance of a 32-GB small form factor HP server SSD with that of a 15K Midline SAS hard disk drive. Sequential read performance is comparable between the two drives, while both random and sequential write performance is slower on the SSD. However, in random read performance the SSD achieved over twelve times the performance of the SAS hard disk drive.

Table 3. Comparison of SSD and HDD performance

	HP server SSD	HP SFF 15K SAS HDD
Random reads (4 KB)	4300 IO/s	340 IO/s
Random writes (4 KB)	100 IO/s	285 IO/s
Sequential reads (64 KB)	100 MB/s	105 MB/s
Sequential writes (64 KB)	80 MB/s	105 MB/s

Reliability of HP server SSDs

Reliability is an important criterion for any storage medium, and it is especially important when considering a storage device that will be used in servers. Because they have no moving parts, there has been a temptation to think that all SSDs should be more reliable than hard disk drives. However, this is not always the case. Good SSD controller designs must successfully manage the basic characteristics of NAND flash arrays while minimizing and correcting problems caused by the basic NAND error modes, including the following:

- Read disturbs
- Program (write) disturbs
- Hot charge injection (bad cells can flip bits)

HP solid state drives for servers employ a variety of mechanisms to deliver a level of reliability that is required in the server environment:

- Using more reliable, longer-lasting SLC NAND technology
- Over-provisioning NAND memory to provide a longer lifecycle
- Wear leveling and block management
- Read and write algorithms that significantly reduce the frequency of NAND error modes

Using these technologies, HP solid state drives for servers are able to achieve a level of reliability equivalent to or slightly greater than current HP Midline disk drives for servers.

Perhaps more importantly for particular applications, HP server SSDs can deliver this level of reliability under conditions that are unsuitable for traditional disk drives, for example in high temperatures and in environments where drives are subject to greater shock and vibration. Table 4 compares the operating envelope of an HP server SSD with that of an HP small form factor SAS enterprise drive.

Table 4. Comparison of SSD and HD operating envelopes

	HP server SSD	HP SFF 15K SAS HDD
Operating temperature	0° – 70° C	10°– 35° C
Operating shock	1500 g (.5 ms half sine wave)	30 g (2 ms half sine wave)
Vibration	20 g peak 10 – 2000 Hz	1.5 g (rms) 10 – 500 Hz
Power consumption (active)	Under 2 watts	8 – 9 watts

SSD futures

The HP server SSDs introduced in late 2008 represent the first generation of SSDs for ProLiant servers. These drives have what will be eventually be considered as entry class capabilities for SSDs. Continuing developments in NAND flash and SSD controller technologies are leading to relatively rapid advances in performance, reliability, and storage capacities for SSDs. In the first half of 2009, HP expects to introduce hot-plug SSDs using standard drive carriers that will be supported across the ProLiant product family. HP is working to develop separate classes of SSDs, each designed to meet the specific capacity, performance, and reliability requirements for particular use environments. Table 5 summarizes the current goals for creating Entry, Midline, and Enterprise SSDs in the future as the technology continues to evolve.

Table 5. Goals for future classes of HP server SSDs

	Entry/Midline	Enterprise
Interface	3 Gb/s SATA	6 Gb/s SAS dual port
NAND memory	SLC based to start MLC as capacity allows	SLC based (for maximum endurance)
Reliability	Optimized for constrained workloads	Optimized for 100% workloads
Latency	< 1 ms for 512-byte random writes	<100 μ s for 512-byte random writes
Maximum capacity	2 times Enterprise SSD	Equal to SFF15K SAS disk drives
Other features	Full memory path error detection Always-on write cache with hot removal protection	Full memory path error detection Always-on write cache with hot removal protection
Power consumption (Active)	Under 2 watts	2 – 9 watts

Summary

Solid state drives using NAND flash memory represent a new and emerging storage product class for use in certain server applications. The first generation of server SSDs provide performance and reliability that are comparable to that of Midline hard disk drives.

For more information

For additional information, refer to the resources listed below.

Resource description	Web address
HP Solid State Storage	www.hp.com/go/solidstate
HP ProLiant drives (including solid state drives)	www.hp.com/products/harddiskdrives
Link to compatibility table for HP server SSDs	http://h18004.www1.hp.com/products/servers/proliantstorage/drives-enclosures/docs/index.html

Call to action

Send comments about this paper to TechCom@HP.com.

© 2008 Hewlett-Packard Development Company, L.P. The information contained herein is subject to change without notice. The only warranties for HP products and services are set forth in the express warranty statements accompanying such products and services. Nothing herein should be construed as constituting an additional warranty. HP shall not be liable for technical or editorial errors or omissions contained herein.

TC081005, October 2008

