Innovative hard disk technologies for enhanced simplicity, performance and availability
October 2004

Innovation and Value in Server Disk Storage

By Mark T. Chapman
IBM Systems and Technology Group

Introduction

As always in this industry, legacy technology gives way to newer technologies offering greater speed, new capabilities, higher availability, and more customer choice.

Perhaps there is no better example of this than in the hard disk drive arena. Early drives were slow and used the PC/AT interface. It didn't take long for the need for faster technology to appear. This resulted in the original SCSI interface. For the next two decades, we had a duality of inexpensive AT-attached (ATA, also called IDE or EIDE) drives defining the low end, and higher-performance, higher-availability SCSI-attached drives defining the high end.

ATA performance gradually increased over many generations: ATA-33, ATA-66, ATA-100, and finally ATA-133. ATA technology has nearly reached its performance limit. SCSI-1, meanwhile, evolved into Fast SCSI, Wide, Fast/Wide, Ultra, Wide Ultra, Ultra2, Wide Ultra2, Ultra3 or Ultra160, and Ultra320, as the need for speed and capability increased. Other differentiators between ATA and SCSI are the availability of hot-swap SCSI drives and SCSI RAID. Along the way, a new player appeared at the extreme high end: fiber-attached drives, capable of external attachment at great distances.

This "separation of powers" existed for two decades, a lifetime in the computer industry. However, today we're seeing another transition in technology—this time from a parallel interface to a smaller serial interface. This may sound trivial, but it brings with it several significant advantages. Furthermore, we are seeing both RAID functionality and swappability features being added to disk drive controllers at the low-end that used to be the exclusive domain of high-end controllers.

The dichotomy of higher-priced, higher-performing, higher function SCSI drives at one end, and lower-priced, lower-performing, lower function ATA drives at the other, still exists today, but with the emergence of SATA and SAS drives, the gap will narrow significantly.

Storage form factor has also seen a number of transitions, first from 5.25-inch full-height to half-height drives, then from 5.25-inch to 3.5-inch, and then 3.5-inch half-height to slim-high drives. Now server density requirements are driving the evolution to 2.5-inch slim form factor drives.

This paper will briefly:

- Explore the differences between parallel ATA and Serial ATA (SATA), and between parallel Ultra320 SCSI and Serial Attached SCSI (SAS); and compare them all to optical fiber-attached (Fibre Channel) drives
- Describe the benefits of "simple-swap" drives as a third alternative to fixed and hot-swap drives
- Describe the benefits of integrated RAID for both SCSI and SATA
- Describe the benefits of the upcoming shift from 3.5-inch to 2.5-inch drives
- Illustrate how IBM @server xSeries® servers are leading the way in 21st century hard disk drive technology.

Commonalities between Serial ATA and Serial Attached SCSI

ATA and SCSI technologies were designed at different times with different goals, and resulted in completely different interfaces for hardware and low-level software. This made life more difficult for vendors who had to design their products to accommodate both storage architectures. Not coincidentally, it made the products more expensive to buy. By contrast, SATA and SAS were designed somewhat in parallel to incorporate many of the same features and offer a degree of hardware and software-level compatibility. In fact, the *SCSI Trade Association* and the *Serial ATA II Working Group* are cooperating to enable compatibility between the two device interface specifications.

For example, both specs use slender four-wire, telephone-like cables for connection, rather than the big bulky ribbon cables required by today's parallel ATA and SCSI products. These thinner cables enhance airflow, and thus system cooling. In addition, the serial interface allows drives to be connected to the system by either cable or direct backplane attachment—something the ATA interface design doesn't permit. In fact, the cable/backplane connectors are *identical* for both SATA and SAS; SAS drives simply use more of the pins than SATA drives do. A system designed with a Serial Attached SCSI interface could support either SAS or SATA drives using the same cables or backplane connectors.

Because both serial architectures use a point-to-point topology, rather than the resource-sharing method required by the older parallel design, the need for complicated jumper configurations (e.g., to differentiate masters from slaves) is eliminated. This simplifies and speeds up the process of deploying or replacing drives.

Perhaps just as importantly, whereas both ATA and SCSI are approaching their performance limits, at 133MBps and 320MBps, respectively, the SATA and SAS architectures are just beginning at approximately those levels. SATA, for example, is designed, over time and in future implementations, to support speeds of 600MBps and higher.

Parallel ATA to Serial ATA

The first transition, from the traditional inexpensive ATA/IDE/EIDE-type drive to Serial ATA, is already under way. It's occurring for a number of reasons: better performance, simplicity of design, ease of use, serviceability, scalability, and lower cost. Because the new IBM @server xSeries 206 (x206) and x306 servers were the first to incorporate SATA technology, we will compare them to their predecessors, the x205 and x305, to illustrate the benefits of SATA drives and controllers. (Other xSeries models, including x226 and x336 servers, as well as the @server 326, also support SATA drives.)

New Capabilities for SATA

As the entry-level *tower* server in the xSeries product line, the x205 incorporated ATA technology. The ATA-100 controller had a maximum disk throughput of **100MB** per second. By comparison, in the **x206** the performance of the SATA controller embedded in the Intel® E7210 chipset is increased by **50%**, to **150MB** per second. System administrators also benefit from the elimination of drive jumpers.

Finally, using a backplane connection rather than signal cables, the x206 offers a new capability called "simple-swap" drives. In the past, ATA drives have been "fixed," or nonremovable without tools. They lacked the convenience and productivity benefit of hot-swappability, which was offered only by high-end SCSI drives in midpriced-or-higher servers.

The **simple-swap SATA** drives implemented in the x206 and x306 are *not* hot-swap drives, but they offer much the same advantage. *Like hot-swap drives, they are tray-mounted and can be plugged in and removed without the need for tools.* However, unlike hot-swap drives, the server must be powered down at the time. This is a great improvement over traditional ATA drives that require not only powering down the system, but also removing the chassis cover, fussing with small screws, and setting jumpers on each drive.

The **x306**, the entry-level *rack-mount* server in the xSeries line, offers the same SATA benefits as the x206, but in a smaller, 1U, package. With up to 42 1U servers in a standard 42U rack, servicing traditional IDE/ATA drives can be a costly inconvenience. Before the drives can be added or replaced, the server first has to be removed from the rack. This entails disconnecting Ethernet and systems management cables, and possibly restructuring cabling configurations. By contrast, the front-accessible simple-swap drives can be replaced easily without the need to

remove the servers from the rack. Simply shut the server down, pull the old drive/tray out, slide the new drive/tray in and power the system back on. No tools or cable reconfiguring are required.

- **Simple-swap drives — ease of use**
- **Toolless customer installation**
- **Improved availability**

RAID Innovations for Entry-Level Servers

RAID, whether used for higher performance, higher-availability, or both, has become a must-have feature of most mid-to-high-end servers. Until recently, RAID controllers were too costly to be used as add-ons for low-cost servers. However, with the advent of *integrated* entry-level RAID capabilities, RAID is affordable for everyone.

All models of the x206 and x306 incorporate both RAID-1 disk mirroring (for high availability) and RAID-0 data striping (for high performance) via the integrated IBM ServeRAID™-7e feature. All that is needed is two internal SATA or SCSI hard disk drives. If additional RAID levels are required, optional ServeRAID controllers can be added. The optional controllers also contain onboard processors and dedicated cache memory, for enhanced RAID performance.

Table 1 summarizes the improvements in storage technology from the x205 to the x206.

Feature	x205 ATA	x206 SATA	x205 SCSI	x206 SCSI
HDD controller throughput	100MBps	**150MBps**	160MBps	**320MBps**
Number / type of drives supported	3 fixed ATA	3 **simple-swap SATA,** *plus* 1 fixed **SATA**	4 fixed Ultra160; or 3 hot-swap Ultra160	4 fixed **Ultra320**; or 3 hot-swap **Ultra320**
HDD capacity	360GB	**500GB**/**1.0TB**[1]	587.6GB fixed; or 220.2GB H/S	**880.8GB fixed**; or **587.2GB H/S**
Integrated RAID levels	N/A	**RAID-0/1**	N/A	**RAID-0/1**

Table 1. *Storage innovations in next-generation x206 servers*

Table 2 summarizes the improvements in storage technology from the x305 to the x306.

Feature	x305 ATA	x306 SATA	x305 SCSI	x306 SCSI
HDD controller throughput	100MBps	**150MBps**	160MBps	**320MBps**
Number / type of drives supported	2 fixed ATA	2 **simple-swap SATA**	2 fixed Ultra160 SCSI	2 fixed **Ultra320** SCSI
HDD capacity	240GB	**500GB**	293.6GB fixed	293.6GB fixed
Integrated RAID levels	N/A	**RAID-0/1**	N/A	**RAID-0/1**

Table 2. *Storage innovations in next-generation x306 servers*

[1] **2** x 250GB drives natively; **4** x 250GB drives when using the optional IBM ServeRAID-7T SATA controller.

In addition to hundreds of megabytes of internal disk capacity, both servers can benefit from *terabytes* of external storage, if needed:

- **x206** — **2TB** (14 x 146.8GB) of **hot-swap Ultra320 SCSI**[2] storage
- **x306** — **2TB** (14 x 146.8GB) of **Fibre Channel**[3] storage

Fiber-Attached Storage

Another innovative drive technology is Fibre Channel (FC). FC drives have been serial-attached since the beginning, and are physically similar to parallel SCSI drives in terms of capacities, rotational speeds, RAID levels and SCSI software layer support. FC commands a price premium over SCSI, both for the drives themselves and for the FC Host Bus Adapters that are required. For this reason, FC drives and controllers are not shipped as internal server drives, but are used for external storage. External storage requires a high degree of availability through redundancy, as well as robust storage management tools.

There are several main advantages to Fibre Channel that more than justify the added cost of FC drives and controllers for high-capacity external storage:

- The fiber optic link allows drive enclosures to be *direct-attached* as far as **ten kilometers** (**6.2 miles**) from the server. This offers great advantages in terms of disaster avoidance and recovery.
- FC SANs are extremely fast, with host bus adapters (HBAs) achieving interface speeds of **2 Gbps** or more between the servers and the storage devices.
- Fibre Channel drives can be incorporated into Storage Area Networks (SANs), supporting up to 16 *million* device addresses. Using specialized storage servers, customers can configure SANs with many terabytes of storage that are accessible to everyone on a network. For example, an IBM **TotalStorage DS4300 Turbo** storage server with seven IBM **TotalStorage DS4000 EXP700** expansion units can support up to **16.4TB** of Fibre Channel storage. (Alternatively, using eight IBM **TotalStorage 4000 EXP100** expansion units, a **DS4300** or a **DS4300 Turbo** storage server can support up to **28TB** of *SATA* storage.) A **TotalStorage DS4400** or **TotalStorage DS4500** storage server and sixteen **EXP700** expansion units increase the FC storage capacity to **32.8TB.** (Using a **DS4500** and sixteen **EXP100** expansion units instead extends the SATA storage capacity to **56TB**.)

> - **High-availability configurations**
> - **High performance**
> - **Geographic dispersal — allows increased distance between server and storage**
> - **Storage Area Network management and scalability**

iSCSI Storage

A recent entry in the SAN field is iSCSI (Internet SCSI). For an organization without a heavy investment in FC infrastructure, and for whom performance is not the primary consideration, iSCSI offers alternative benefits for most xSeries servers:

- Little or no additional training is needed to add IP SAN skills to existing IP network skills. Not only does this save a lot of money on training, it means your people can get up to speed on SAN support sooner.

---

[2] Requires a ServeRAID-6M controller and one IBM TotalStorage 4000 EXP400 Storage Expansion Unit.

[3] Requires an IBM TotalStorage 4000 Host Bus Adapter and either a TotalStorage 4000 EXP700 Storage Expansion Unit or a TotalStorage DS400 Storage Server.

- With iSCSI you can use your existing CAT5 cabling and Ethernet switches, routers and hubs. At the low end, standard Ethernet NICs are far less expensive than FC HBAs. Also, existing NAS servers can easily be incorporated into an iSCSI SAN. This allows you not only to save on additional SAN hardware, but to aid in a SAN consolidation effort. If you already have an FC SAN installation, it can interconnect with an iSCSI SAN through an IP storage bridge or gateway (which converts the FC protocol to and from iSCSI).
- For some customers, the ability of iSCSI SANs to span the entire extent of an IP network may be its biggest advantage. It means that there is no practical limit to how far away offsite backup can be. The remote sites can also be used for clustering and mirroring/replication. Organizations can even elect to link to storage service providers for storage-on-demand or other applications.
- Using specialized storage servers, customers can configure SANs with many terabytes of storage that are accessible to everyone on an IP network at an extremely competitive price. For example, an IBM **TotalStorage DS300** storage server provides up to **2TB** of iSCSI storage, using inexpensive Ultra320 SCSI drives. As many DS300 units as desired can be SAN-attached using standard Gigabit Ethernet Adapters or iSCSI host bus adapters.

- **Short learning curve**
- **Standard and familiar components**
- **IP networkwide extent**
- **Terabytes of inexpensive storage**

New Form Factor Drives

Another transition already occurring is to decrease the size of drives from the current 3.5-inch to a 2.5-inch form factor. In one sense there is nothing new here: notebook computers have used 2.5-inch drives for years. What *is* new is that some of these small form-factor drives have now reached the reliability levels needed for heavy-duty enterprise server utilization. In fact, IBM BladeCenter HS20 blade servers have been using 2.5-inch drives since late 2002.

This migration offers server users a number of advantages. Comparing typical 3.5-inch and 2.5-inch 73.4GB 10K RPM Ultra320 SCSI drives, the 2.5-inch models:

- *Take **70%** less space, by volume*, allowing either more drives to fit in the same space, or for servers to be made smaller while including the same number of drives
- *Weigh nearly **70%** less* (more than a pound per drive less), reducing the strain on datacenter floors and making full racks easier to move, if needed
- *Draw **40%** less power* and *produce that much less heat*, which can help lower data center electrical operating expenses
- *Have **20%** greater non-operating shock ratings*
- *Have **17%** greater mean time between failure (MTBF) ratings*, due to lower power draw and less heat
- *Have **15%** faster seek times* (due to the smaller disk diameters)

IBM was the first Tier 1 vendor to ship 2.5-inch SCSI drives in a traditional (non-blade) server in 2004, via the x336. SAS drives in a 2.5-inch form factor are expected in 2005. The downside of the migration to 2.5-inch drives is that *initially* the capacities of the drives will be lower, drive rotational speeds will be lower (no 15K RPM drives at first; only 10K), and the prices will be somewhat higher. It is anticipated that all of these shortcomings will be overcome by the second generation of drives.

- **Increased drive density via space efficiency**
- **Reduced system power requirements**
- **Positioned for select rack-dense usage**

Choosing between SATA, SCSI and FC

You can use the following guidelines to choose between parallel ATA, SCSI and FC:

- For good performance (up to **150MBps**; 7200 RPMs) at a low-cost, with basic RAID levels, go with **SATA**.
- For higher performance (up to **200MBps**; 10K-15K RPMs), hot-swappability, high MTBF rates, and additional RAID levels, at distances of up to **10km**, go with **Fibre Channel**.
- For the highest performance (up to **320MBps**; 10K-15K RPMs), hot-swappability, high MTBF rates, and additional RAID levels, either internally or nearby, go with **SCSI**.

Parallel SCSI to Serial Attached SCSI

The migration from SCSI to SAS will occur a bit later than the one from ATA to SATA. SAS-equipped systems are expected to begin arriving in late 2004 or early 2005.

SAS, like parallel SCSI before it, was designed primarily as a storage architecture for servers, with higher levels of performance and availability built in, and at a higher price tag than parallel or serial SATA. Like SCSI, SAS drives can be used inside a server, or in a local external expansion unit or Network Attached Storage (NAS) server, and SAS is compatible with existing SCSI applications and middleware layers. Like parallel SCSI, SAS drives will offer support for hot-swappability and multiple RAID levels. Due to the inherent advantages of SAS drives, they are expected to quickly supplant SCSI in all roles.

New Capabilities for SAS

As with SATA drives, SAS drives benefit from the lack of jumpers and the thinner, more flexible signal cables. The same SAS controller can support *either* SAS or SATA drives—*or both*—in the same system. Imagine trying to attach ATA drives to an Ultra320 SCSI cable! The use of a single cable or backplane part for both types of drives means a simpler design and lower costs for the system manufacturer, and thus for the customer.

Like Ethernet, SAS achieves great scalability by using a switched point-to-point topology. Where SCSI is limited to **15** devices per port, extended drive addressing allows SAS to support up to **4,032** devices per port. Low-cost “expanders” provide additional ports, for a theoretical scalability of up to **16,000** drives per server in the future.

SAS also supports *dual-porting*. Dual ports provide two separate data paths to the drive, for higher performance and elimination of a single point of failure. This makes SAS a superior technology for enterprise storage systems.

Parallel SCSI is already approaching the practical limits of performance at **320MBps**. By contrast, the first generation of SAS drives will begin at **300MBps**. The SAS roadmap calls for performance increases to as much as **1,200MBps** (1.2GBps) in succeeding generations.

- **Increased performance and throughput**
- **Cabling simplification and thermal improvements**
- **Robust technology for demanding server environments**

Summary

In the past, entry-level servers were typically minimal configurations, especially in regards to the storage subsystem, with small capacities, slow throughput, and limited availability features. Today, IBM has taken capabilities that used to be reserved for only middle- and high-end servers and offered them in servers that are entry-priced.

It's not merely a matter of evolution, of waiting for the cost of current technology to drop enough that it is affordable to use in inexpensive servers, such as Ultra320 SCSI and integrated RAID. It's also a matter of revolution—of using brand new technologies that improve the availability, serviceability, and performance of those servers, including SATA drives and interchangeable backplanes for both SATA and SAS. More than that, it's a matter of *innovation*—taking industry-standard technologies such as SATA and making them even better by offering toolless simple-swap capability to ease servicing and help reduce system downtime; or providing the industry's best systems management software and hardware on even our entry-level servers.

Most of all, it's a matter of meeting customer demand for reliability, availability, serviceability, manageability, performance, and of course, flexibility—all at an entry-level price. IBM offers a choice of tower or rack-optimized servers; a choice of Intel Celeron® or Intel Pentium® 4 processor; a choice of simple-swap SATA, fixed SCSI, or hot-swap SCSI; a choice of built-in RAID or higher-performance optional RAID controllers; a choice of built-in systems management features or an even more powerful optional systems management controller; a choice of Microsoft® Windows® and Linux® operating systems; and soon a choice between SCSI and SAS storage and between 3.5-inch and 2.5-inch drives.

Whether you require only minimal internal storage or a highly scalable storage area network, IBM can provide you with a wide array of choices, and the management tools and services you need.

Additional Information

For more information on IBM @server xSeries directions, products and services, visit our Web site at http://**ibm.com**/eserver/xseries. From the xSeries home page, select **Library** and you will see links to the different types of documentation available.

For more information about IBM FAStT products, refer to http://storage.**ibm.com**/disk/fastt/index.html.

For more on other SAN, NAS, and remote storage products: http://storage.**ibm.com**/snetwork

Go to *http://www.scsita.org* for more information about the SCSI Trade Association.

See *http://www.serialata.org* for more information about the Serial ATA II Working Group.

IBM Systems and Technology Group
Dept. U2SA
3039 Cornwallis Road
Research Triangle Park, NC 27709

Produced in the USA
10-04

Visit http://**ibm.com**/pc/safecomputing periodically for the latest information on safe and effective computing. Warranty Information: For a copy of applicable product warranties, write to: Warranty Information, P.O. Box 12195, RTP, NC 27709, Attn: Dept. JDJA/B203. IBM makes no representation or warranty regarding third-party products or services including those designated as ServerProven or ClusterProven. Telephone support may be subject to additional charges. For onsite labor, IBM will attempt to diagnose and resolve the problem remotely before sending a technician.

IBM, the IBM logo, the e-business logo, ServeRAID and xSeries are trademarks of IBM Corporation in the United States and/or other countries. For a list of additional IBM trademarks, please see http://**ibm.com**/legal/copytrade.shtml.

Intel, Intel Celeron, Intel Pentium, Intel Inside and the Intel Inside logo, are trademarks or registered trademarks of Intel Corporation or its subsidiaries in the United States and other countries. Linux is a trademark of Linus Torvalds. Microsoft, Windows, and the Windows logo are trademarks or registered trademarks of Microsoft Corporation.

Other company, product and service names may be trademarks or service marks of others.

All statements regarding IBM's future direction and intent are subject to change or withdrawal without notice, and represent goals and objectives only. Contact your local IBM office or IBM authorized reseller for the full text of a specific Statement of General Direction.

IBM reserves the right to change specifications or other product information without notice. References in this publication to IBM products or services do not imply that IBM intends to make them available in all countries in which IBM operates. IBM PROVIDES THIS PUBLICATION “AS IS” WITHOUT WARRANTY OF ANY KIND, EITHER EXPRESS OR IMPLIED, INCLUDING THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE. Some jurisdictions do not allow disclaimer of express or implied warranties in certain transactions; therefore, this statement may not apply to you.

This publication may contain links to third party sites that are not under the control of or maintained by IBM. Access to any such third party site is at the user's own risk and IBM is not responsible for the accuracy or reliability of any information, data, opinions, advice or statements made on these sites. IBM provides these links merely as a convenience and the inclusion of such links does not imply an endorsement.

Information in this presentation concerning non-IBM products was obtained from the suppliers of these products, published announcement material or other publicly available sources. IBM has not tested these products and cannot confirm the accuracy of performance, compatibility or any other claims related to non-IBM products. Questions on the capabilities of non-IBM products should be addressed to the suppliers of those products.

MB, GB, and TB = 1,000,000, 1,000,000,000 and 1,000,000,000,000 bytes, respectively, when referring to storage capacity. Accessible capacity is less; up to 3GB is used in service partition. Actual storage capacity will vary based upon many factors and may be less than stated.

Maximum internal hard disk and memory capacities may require the replacement of any standard hard drives and/or memory and the population of all hard disk bays and memory slots with the largest currently supported drives available.

Many factors determine the sizing requirements and performance of a systems architecture. IBM assumes no liability for the methodology used for determining the configurations recommended in this document nor for the results it provides. Any performance data contained herein was determined in a controlled environment. Therefore, the results obtained in other operating environments may vary significantly. Actual results may vary. Users of this presentation should verify the applicable data for their specific environment.