White Paper

An Introduction to EMC CLARiiON Storage Device Technology

Applied Technology

Abstract

This white paper compares and contrasts the merits of four types of storage devices that can be used on EMC® CLARiiON®: enterprise Flash drive (EFD), Fibre Channel (FC), Serial Attach SCSI (SAS), and Serial ATA (SATA).

June 2009

Part Number H4208.2

Table of Contents

Executive summary

There are four types of storage device technologies available for use on EMC® CLARiiON®: enterprise Flash drive (EFD), Fibre Channel (FC), Serial Attach SCSI (SAS), and Serial ATA (SATA). While all of these drives store data, they have different operating characteristics, interfaces, and price points that make them more suitable to different operating environments. For example FC, SAS, and SATA drives have spinning media whereas enterprise Flash drives have solid state media – this characteristic makes them suitable for different kind of environments.

There are two main types of operating environments: performance-intensive environments and capacity-intensive environments. Performance environments need high throughput, high bandwidth, and high reliability. Typical performance environments are transaction-based systems with workloads characterized by random reads and writes.

Capacity-intensive environments need large amounts of low-cost storage, high reliability, and modest throughput and bandwidth performance. Typical capacity-intensive environments are archival systems with a workload characterized by sequential reads.

EFDs are recommended for performance-intensive applications or parts of applications with low-response-time and high-throughput requirements. FC hard drives should continue to be the choice for environments with large-capacity high-performance requirements. SAS hard drives provide performance and reliability that are equivalent to FC drives. SATA hard drives are the choice in modest-performance and high-capacity environments. In addition, SATA drives can provide energy-efficient bulk storage capacity at low cost.

However, the choice between performance and capacity is not always clear. Many CLARiiON systems operate in "mixed" environments that have a primary mission to provide storage for a performance workload but must also provide storage for a capacity-intensive workload. In this case, a blend of storage device types (EFD, FC, and SATA on CX arrays or SAS and SATA on AX arrays) may be the most appropriate and cost-effective solution. In addition, there are CLARiiON systems that operate in environments with requirements for *both* performance and capacity. In this case, a high-capacity "EFD and FC" or "all SAS" drive solution is merited. Choice of a drive type also depends on the type of CLARiiON storage system. CLARiiON CX arrays support EFD, FC, and SATA drive types while CLARiiON AX arrays support SAS and SATA drives only.

Introduction

This white paper compares and contrasts the merits of the different drive types (EFD, FC, SAS, and SATA) that can be used on the CLARiiON. This white paper was formerly titled *An Introduction to EMC CLARiiON Hard Drive Technology*.

With more open-system storage capacity shipping this year than in all past years combined, a new standard for storage system performance is required to access all this data in an efficient manner. With its CLARiiON storage systems, EMC steps up to meet this challenge with its lineup of some of the fastest and highest capacity storage devices available on the market today. The 1 TB 5,400/7,200 rpm SATA II drives, 400 GB 10k rpm Fibre Channel/SAS drives, 146/300/450 GB 15k rpm Fibre Channel/SAS drives, and 73/400 GB enterprise Flash drives all redefine storage-system capacities and performance.

Audience

This white paper is intended for system engineers, EMC partners, members of EMC and partner sales and professional services, and engineers wanting an introduction on the types of storage devices and their characteristics available on the EMC CLARiiON.

Terminology

Average seek time — The average amount of time it takes for the head stack assembly (read/write heads) to get from one point on the media surface (*A*) to another point on the same surface (*B*). Average seek time is obtained by averaging the seek times for all random seeks over a certain period of time. Naturally, this time is reduced as the spindle motor speed is increased because data becomes available more quickly. Seek times are measured separately for read and write operations. Read operations take less time than write operations.

Flash memory – Flash memory is a non-volatile memory that can be electrically erased and reprogrammed. It is a specific type of electrically erasable programmable read-only memory (EEPROM) that is erased and programmed in blocks. It does not need any power to maintain the information stored in the chip. The individual cells used for storing bits of information can be of NOR or NAND architecture.

Fibre Channel — A high-bandwidth data-transfer protocol that uses optical or copper cables to connect devices and is widely used to connect servers to high-performance storage systems in a storage area network (SAN).

Magic 8 bytes – All CLARiiON systems use a disk sector format of 520 bytes. This format allows the storage of 512 bytes of data, as in most disk systems, but also provides 8 bytes per sector for metadata, which CLARiiON algorithms use for data integrity checks. These additional bytes include a linear checksum, write stamp, parity shed stamp, and time stamp. The RAID protection level of the sector determines which of these stamps is used by CLARiiON.

Rotational latency — The disk platters must spin to move the data on a given track under the read/write heads. This takes time. On average, the platter must rotate half of a revolution to reach the sector. This latency time is inversely proportional to the disk rotation speed and is reduced as the spindle motor speed increases. The average rotational latency specification of a 15,000 rpm drive is 2.0 ms, which is the time for a half revolution of the disk surface.

Serial Advanced Technology Attachment (SATA) — An enhancement that uses thinner cables and provides better performance over parallel drive technology. SATA II includes improvements such as Native Command Queuing (NCQ) to improve performance and reliability.

Serial Attach SCSI (SAS) — An I/O technology that uses a serial point-to-point interface instead of a parallel bus interface, as with parallel SCSI.

Spindle motor speed — The actual speed of the spindle motor assembly measured in rpm or revolutions per minute. One complete revolution is the amount of time it takes to make one full revolution from a given start point on the disk surface (*Index*) back to the given start point Index. The amount of time it takes on a 15,000 rpm disk drive to make one full revolution is 4 ms, compared to approximately 6 ms on a 10,000 rpm disk drive. This is a roughly 33 percent decrease in the amount of time it takes to make one full revolution of the disk media surface.

Transfer rate — The rate at which data is transferred from either the drive to the target host (external) or from the media surface to the head assembly (internal). Although the internal transfer rate of the hard disk increases as the spindle speed increases (head-to-media transfer rate), the external transfer rate typically stays the same due to the interface chipset architecture of the hard disk drive itself. Unless a higher transfer rate interface chipset is used on the hard disk drive itself, the external transfer rate will stay the same as that of the predecessor drive.

Enterprise Flash drive/solid state technology

EFDs, commonly known in the IT world as solid state drives (SSDs), that are used in CLARiiON CX4 arrays differ significantly from the solid state technology used in consumer electronics, particularly in their performance and reliability characteristics. Consumer electronics are built around commodity class technology where price is the primary consideration. These consumer grade devices do not have the performance and resiliency that enterprise applications require. EFDs address the performance and reliability challenges that are associated with Flash technology. An EMC enterprise Flash drive is a sophisticated system in itself that has been designed to EMC's exacting engineering requirements and specifications. EFDs have dual-ported front-end controller interfaces that allow the drive to deliver data over two different channels; a highly protected and integrated processing engine; a large amount of DRAM to act as a buffer between the array and the relatively slower Flash media; and the ability to simultaneously process multiple read and write I/Os to the Flash media in the back end. The drive is also protected end to end with error detection/correction technology to detect and correct bit errors that may occur. There is a high level of redundancy and error correction built into each drive to make sure that data corruption is always detected and corrected.

EFDs are especially well suited for low-latency applications that require consistent, low (less than 1 ms) read/write response times. The greatest improvements will be seen with highly concurrent, random read workloads, since there is no rotational or seek latency in EFDs. The elimination of mechanical overhead and data placement latency greatly improves application performance and efficiency. All of these factors contribute to the EFDs on CLARiiON being able to deliver very high IOPS with very low response times.

Besides the industry standard techniques used in all CLARiiON storage devices (like DRAM memory between the array and the storage media; a backup power circuit for the DRAM memory; error detection and correction; and bad block management), there are certain features that are specific to EFDs. These features are discussed in the following sections.

Wear leveling

Individual Flash cells support a finite number of erase-write cycles. Utilizing this technology in enterprise storage requires features that improve the drive's reliability. EFDs have more raw NAND Flash capacity internally than the usable amount visible to the CLARiiON array. EMC EFDs balance erase and rewrite operations across their entire raw capacity to sustain maximum bandwidth and reliability, and to optimize the usable life of the drive. EFDs employ this wear-leveling technique to ensure that all cells in the drive are used evenly, and to minimize the risk of "wear-out" to which other industry Flash drive cells are susceptible. This wear-leveling is performed separately for each I/O channel within the EFD. There are two types of wear-leveling:

- Dynamic wear leveling - Pointers to all available blocks within an EFD are maintained in a *free list.* When the number of blocks in the free list drops below a minimum threshold, a background process frees up blocks with the least amount of data by consolidating their data into a single block – similar to the de-fragmentation process in many operating systems. The blocks that are freed up are erased and added to the free list. Since any new block is always allocated from the free list, the drive firmware ensures that the wear is distributed evenly across all available Flash cells in the drive.
- Static wear leveling - EFD firmware maintains a list of how often blocks have been erased, which is a measure of how "worn" a block is. With wear-leveling, the contents of blocks with high erase counts are moved to blocks with low erase counts. The drive firmware keeps the most erased block and least erased block on an I/O channel within a set number of erase cycles of each other.

Multiple I/O channels

EFDs have multiple internal independent I/O channels, which are also referred to as *threads.* This is like having 16 independent service centers for the incoming I/O requests. This allows EFDs to service multiple requests at the same time, unlike traditional spinning drives that can service only one request at a time. This ability greatly enhances the performance of applications with higher concurrency (or parallelism).

Flash cell architecture

In general, there are three major technologies used in all Flash drives:

- NOR-based Flash cell
- Single-Level Cell (SLC) NAND-based Flash cell
- Multi-Level Cell (MLC) NAND-based Flash cell

NOR Flash technology is primarily used where high-density memory and high performance are not the main requirements. It is best suited for code storage and execution – usually in low densities. This makes it ideal for chips that store program code that is rarely updated, such as ROM BIOS code in a motherboard.

NAND Flash cells have a very compact architecture; their cell size is almost half the size of a comparable NOR cell. This characteristic, when combined with a simpler production process, enables a NAND cell to offer higher densities with more memory on a given semiconductor die size. This results in a lower cost per gigabyte. With smaller, more precise manufacturing processes being used in the future, their price is expected to fall even further. A NAND Flash cell also has faster erase and write times compared to NOR-based Flash cells, thus providing improved performance. It has higher endurance limits for each cell, a feature that provides the reliability required for enterprise-class applications.

Flash storage devices store information in a collection of Flash cells made from floating gate transistors. Single-Level Cell (SLC) devices store only one bit of information in each cell (binary), whereas Multi-Level Cell (MLC) devices store more than one bit per Flash cell by choosing between multiple levels of electrical charge to apply to its floating gates in the transistors (Figure 1).

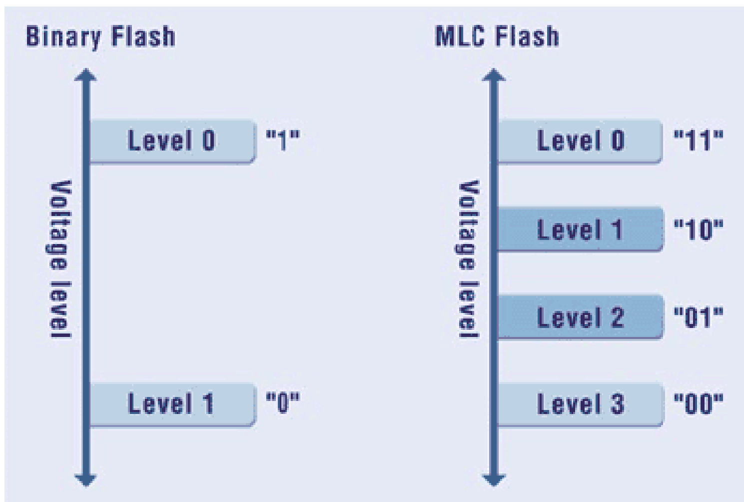

Figure 1. Comparison between SLC and MLC Flash cell data storage[1]

Since each cell in MLC Flash has more information bits, an MLC Flash-based storage device offers increased storage density compared to a SLC Flash-based version. The downside is that MLC NAND has lower performance and reliability because of its inherent architectural limitations. Higher functionality further complicates the use of MLC NAND, making it necessary to implement more advanced Flash management algorithms and controllers. SLC and MLC NAND offer capabilities that serve two very different types of applications – those requiring high performance at an attractive cost per bit (MLC) and those seeking even higher performance over time that are less cost sensitive.

Taking into account the kind of I/O profiles in enterprise applications and their requirements, EMC enterprise Flash drives have the SLC NAND Flash architecture. Table 1 compares the SLC and MLC Flash characteristics.

[1] Kaplan, Francois, "Flash Memory Moves from Niche to Mainstream ," *Chip Design Magazine*, April/May 2006

Table 1. SLC and MLC Flash comparison

Features	MLC	SLC
Bits per cell	2	1
Endurance (ERASE/WRITE cycles)	<10K	<100K
Read service time (Max)	50µs	25µs
Write service time (Typical)	600–900µs	200–300µs
Block Erase (Typical)	2 ms	1.5–2 ms

Although SLC NAND offers a lower density, it also provides an enhanced level of performance in the form of faster reads and writes. Because SLC NAND stores only one bit per cell, the likelihood for error is reduced. SLC also allows for higher write/erase cycle endurance, making it a better fit for use in applications requiring higher reliability, and increased endurance and viability in multi-year product life cycles.

Hard drive technology

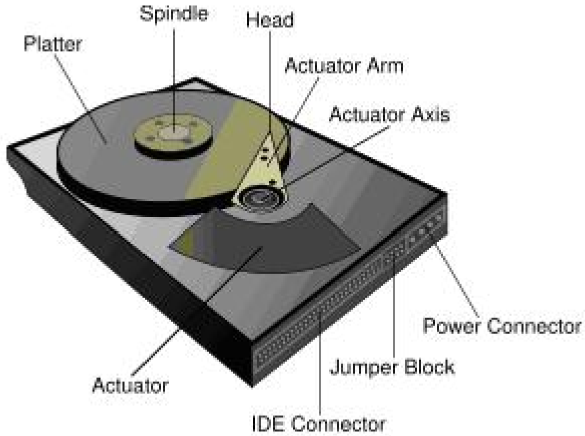

Figure 2. Physical drive

Hard disk drives (HDDs) record data by magnetizing a ferromagnetic material directionally, to represent either a 0 or a 1 binary digit. The drives read the data back by detecting the magnetization of the material. A typical HDD design consists of a spindle that holds one or more flat circular disks called platters, onto which the data is recorded. The platters are made from a non-magnetic material, usually glass or aluminum, and are coated with a thin layer of magnetic material. Older disks used iron oxide as the magnetic material, but current disks use a cobalt-based alloy.

The platters are spun at very high speeds. Information is written to a platter as it rotates past the read-write heads that operate very close to the magnetic surface. The read-write head is used to detect and modify the magnetization of the material immediately under it. There is one head for each magnetic platter surface on the spindle, mounted on a common arm. An actuator arm (or access arm) moves the heads on an arc (roughly radially) across the platters as they spin, allowing each head to access almost the entire surface of the platter as it spins. The arm is moved using a voice coil actuator or (in older designs) a stepper motor.

The magnetic surface of each platter is divided into many small sub-micrometer size magnetic regions, each of which is used to encode a single binary unit of information. In today's HDDs each of these magnetic regions is composed of a few hundred magnetic grains. Each magnetic region forms a magnetic dipole that generates a highly localized magnetic field nearby. The write head magnetizes a magnetic region by generating a strong local magnetic field nearby. Early HDDs used an electromagnet both to generate this field and to read the data by using electromagnetic induction. Later versions of inductive heads included Metal in Gap (MIG) heads and thin film heads. In today's heads, the read and write elements are separate but in close proximity on the head portion of an actuator arm. The read element is typically magneto-resistive while the write element is typically thin-film inductive.

In modern drives, the small size of the magnetic regions creates the danger that their magnetic state will be lost because of thermal effects. To counter this, the platters are coated with two parallel magnetic layers, separated by a 3-atom-thick layer of the non-magnetic element ruthenium, and the two layers are magnetized in opposite orientation, thus reinforcing each other. Another technology used to overcome thermal effects to allow greater recording densities is perpendicular recording, which has been used in some hard drives as of 2006.

Hard disk drives are sealed to prevent dust and other sources of contamination from interfering with the operation of the hard disks heads. The hard drives are not air tight, but rather utilize an extremely fine air filter, to allow for air inside the hard drive enclosure. The spinning of the disks causes the air to circulate, forcing any particulates to become trapped on the filter. The same air currents also act as a gas bearing, which enables the heads to float on a cushion of air above the surfaces of the disks.

The logical drive

Logical drives are created by allocating an entire (physical) drive, part of one drive, or parts of more than one drive. Logical drives are called logical volumes, LUNs, and in some cases virtual disks.

The disk is described as *logical* because it does not actually exist as a single physical entity in its own right. Logical disks are also defined at various levels in the *storage infrastructure stack*. The levels, from top to bottom, are:

- Operating system: Defines *partitions* on the disks to which it has visibility (these disks may be logical themselves).
- SAN: If the SAN is virtualized, a device in the same SAN presents *logical drives* to the host operating systems.
- Storage subsystem: Usually providing some form of RAID where *logical drives* (partitions) are presented to the SAN from the RAID arrays themselves. The RAID arrays actually contain the physical disks. Depending on the RAID type and the configuration used, a set of physical drives form a logical drive. This is applicable both for EFDs and rotating disk drives.

Enterprise Flash drives

EMC CLARiiON 73 and 400 GB (4 Gb/s) EFDs

With this capability, EMC CLARiiON has a *Tier 0* ultra-performance storage tier that transcends the limitations previously imposed by magnetic disk drives. For years, the most demanding enterprise applications have been limited by the performance of magnetic rotating disk media. Tier 1 performance in storage systems has been unable to surpass the physical limitations of rotating disk drives. With EMC's addition of EFDs to CLARiiON CX4, organizations can now take advantage of the drive's ultra-high performance that is optimized for the highest level enterprise requirements.

EFDs contain no moving parts and appear as standard Fibre Channel drives to existing CLARiiON management tools, allowing administrators to manage Tier 0 without special processes or custom tools. Tier 0 Flash storage is ideally suited for applications with high transaction rates and those requiring the fastest possible retrieval and storage of data. Additionally, because there are no mechanical components, EFDs require up to 98 percent less energy per I/O than traditional disk drives.

Until now, enterprises with high performance requirements had limited choices. They could either take a costly approach of spreading workloads over dozens or hundreds of underutilized disk drives, or they could purchase separate and expensive servers and memory storage that essentially add complexity and create storage islands. Now, using EFDs, Tier 0 applications can be closely coupled with other storage tiers within CLARiiON for consistency and efficiency, eliminating the need for time invested in manual data layout or end-of-day data transfers from separate RAM disk or specialized memory storage systems.

Performance of 73 and 400 GB EFDs

Enterprise Flash drives do not contain any moving parts within the drive. This, combined with the multiple parallel I/O channels on the back end, allows EFDs to show performance characteristics that are multiples above a traditional rotating drive. EFDs show the highest improvement in performance (compared to the best performing rotating drives) when the drive workload has a high percentage of small, random read operations and the application is multi-threaded.

The typical expectation for a 15k rpm FC device is about 180 IOPS; the typical expectation for an EFD device is about 2,500 IOPS. Note that these throughput values are measured at the drives; when measuring throughputs at the host/server, parity calculations should also be included. In carefully tuned multi-threaded random small-block environments, EFDs can deliver much lower response times and higher throughput than rotating drives. Random-read I/Os cannot usually be serviced by read-ahead algorithms on the drive or by read cache on the storage system. Therefore, the latency of a random read operation is directly related to the seek time of a disk drive. For rotating drives, this is the mechanical movement of the disk head while it reads the desired area. EFDs pay no penalty for retrieving I/O that is stored in more than one area; as a result their response time is an order of magnitude faster than the response time of rotating drives. However, production workloads that currently run on FC drives should not necessarily be expected to scale at the same multiplier as drive IOPS since replacing the disk drive technology does not change the application thread count or host/SAN service times.

Like their rotating counterparts, EFDs have a wide range of operating points that depend on the nature of the workload that they service. For large block I/Os (64 KB and higher), EFDs tend to use all internal I/O channels in parallel. Since single-threaded sequential I/O streams on FC drives do not suffer seek and rotational latencies (because of the array cache), single-threaded large-block sequential I/O streams will not show major performance improvements with EFDs over FC drives. However, with increased application concurrency (as more threads are added), the load starts to resemble a large block-random workload. In this case, seek and rotational latencies are introduced that decrease FC drive effectiveness but do not decrease EFD effectiveness.

Applications using 73 and 400 GB EFDs

Some applications are better suited to take full advantage of EFD capabilities than others. Applications with these traits are ideal for EFD deployment:

- A minimum small block read response time
- The ability to scale throughput linearly over many concurrent threads
- Avoidance of forced flushing.

Detailed performance guidance for using enterprise Flash drives can be found in *EMC CLARiiON Performance and Availability: Release 28.5 Firmware Update – Applied Best Practices* available on EMC Powerlink®.

Based on the performance considerations mentioned in the previous sections, some general areas of interest for EFD deployment are as follows:

- Hot database tables that require very fast read response time and/or high throughput
- Database temp area

- High write bandwidth applications, like DB load and system backup, that are known to cause forced flushing
- Small, highly active file systems that have a response time SLA
- Metadata control areas in clustered file systems that require minimum read response time (and not necessarily high throughput)

Fibre Channel hard drives

EMC CLARiiON 146, 300, and 450 GB 15k rpm (4 Gb/s) FC drives

The CLARiiON 146, 300, and 450 GB 15k rpm 4 Gb/s hard disk drives dramatically increase performance through improvements in disk operations such as rotational latency and seek rates—the factors that most directly affect access times. These disk drives deliver 3.5 ms average read seek times and 2.0 ms average rotational latency times—both the fastest available in today's hard disk drive market.

Details about other capacity 15k rpm (4 Gb/s) FC drives supported on CLARiiON systems can be found in the *CX4 Series Storage Systems Disk and FLARE OE Matrix* available on EMC Powerlink.

The following section illustrates how the increase in spindle motor rotation speed can make a difference in overall I/O performance when compared to slower rotational-speed hard disk drives.

Performance of 15k rpm disk drives

Analysis of 15k rpm disk drive performance shows that not much of a performance benefit is realized over slower spindle (for example, 10k rpm) drives when running applications that are sequential, read I/O intensive. This is due to the fact that when running sequential I/O, there is virtually no head carriage movement, and spindle speed has little effect in the overall access to sequential data coming from the media surface. Even with larger block sizes in the equation, there is no discernible difference in the overall transfer rates between 10k and 15k rpm disk drives for purely sequential workloads.

The overall performance difference between a 15k rpm and 10k rpm disk drive increases when I/O types are varied, such as random read and random write operations. Larger performance improvements are realized when running the 15k rpm disk drives in random I/O environments. Drives of 15k rpm can yield up to a 35 percent performance improvement in a random environment when compared to the 10k rpm drives.

Capacity does not have an appreciable effect on the performance of disk drives. For example, there is not much performance difference between a 146 GB and a 300 GB 15k rpm FC drive.

Applications using 146, 300, and 450 GB 15k rpm FC drives

Using the 15k rpm drives in applications that use small block, random I/O is an important factor in realizing higher performance benefits. These applications have a tendency to minimize any caching advantages of the storage system. In addition, with applications such as these, the physical access to data on the disk has the greatest effect on overall performance. These small block, random I/O applications reap the greatest benefits from storage-system performance improvements achieved through the new drive’s improved seek and rotational latency times.

Some of the more popular types of applications for the 15k rpm drives include:

- OLTP
- E-commerce
- ERP
- Database
- Web server

- E-mail
- Data replication

EMC CLARiiON 400 GB 10k rpm (4 Gb/s) FC drives

The CLARiiON 400 GB 10k rpm 4 Gb/s hard disk drives dramatically increase capacity and lower power consumption over their 2 Gb/s counterparts through improvements in disk operations such as spindle motor technology, interface chip sets, and disk platter densities—the factors that most directly affect storage capacities and power consumption. These disk drives deliver 3.9 ms average read seek times and 2.98 ms average rotational latency times.

Details about other capacity 10k rpm (4 Gb/s) FC drives supported on CLARiiON systems can be found in the *CX4 Series Storage Systems Disk and FLARE OE Matrix* available on EMC Powerlink.

EMC CLARiiON CX4 series arrays are not sold with 2 Gb/s drives. However, they are still supported in CX4 and legacy arrays as an upgrade. Details about 10k rpm (2 Gb/s) Fibre Channel drives supported on CLARiiON systems can be found in the *CX4 Series Storage Systems Disk and FLARE OE Matrix* available on EMC Powerlink.

Performance of 10k rpm disk drives

When comparing the *same* number of spindles in a random read/write environment, the performance of hard disk drives with different capacities is virtually equal. This is because their performance specifications are almost identical. Replacing 146 GB drives with an equal number of 400 GB drives, without increasing the amount of data, greatly improves performance because of the reduced seek distances. This advantage is reduced as new data is added to the configuration. If 146 GB drives are replaced by a lesser number of 400 GB drives, the performance is reduced, since the throughput of the system is directly proportional to the number of spindles.

The performance of the 146 GB and 400 GB Fibre Channel hard disk drives are similar when applications are sequential read and write I/O intensive. When running sequential I/O, there is virtually little or no head carriage movement, and the overall increase in BPI helps to maintain similar performance characteristics of the 146 GB and the 400 GB disk drives. Even with larger block sizes, there is still no discernible difference in the overall transfer rates between 146 and 400 GB 10k rpm disk drives.

To summarize, when using the 400 GB drives in the right environment, there is up to a 10 to 12 percent performance improvement over the 146 GB previous-generation disk drives. In a sequential workload environment, the performance implications of replacing 146 GB drives with 400 GB ones are different. Seek times are insignificant, so there is no advantage to having more space. The spindle count change also has a lesser impact on throughput because it takes fewer drives to limit system performance. In many cases, we can retain adequate throughput performance in a bandwidth environment by using fewer spindles.

Applications using 400 GB 10k rpm disk drives

To realize cost and capacity benefits, it is important to use the new 400 GB 10k rpm drives in applications suited to higher-capacity environments. Sequential access applications have a tendency to maximize any caching advantages of the storage system, and take advantage of the *increased internal transfer rate* of the drive. In these applications, the speed at which the drive can transfer data from the platter has the greatest effect on overall performance. Thus, medium-to-large block and sequential I/O applications reap the greatest benefits from the new drive’s improved internal transfer rates and higher bit densities.

Some of the more popular types of applications for the 400 GB 10k rpm drives are:

- Database environments
- Online backup for Internet services
- Near-line storage or tape replacement
- Oil and gas exploration

- Life sciences
- Digital A/V and digital editing
- Medical imaging
- Image archival
- Document management
- Data warehousing and data mining

The customer benefits from 400 GB disk drive technology in these areas:

- Price per megabyte
- Reduced footprint
- Reduced power and cooling requirements
- Improved MTBF rates through drive consolidation
- Available on the current CX series product lines

Serial Attach SCSI hard drives

Serial Attach SCSI (SAS) is a computer bus technology primarily designed to transfer data to and from devices like hard drives, CD-ROM drives, and so forth. SAS is a serial communication protocol for direct attached storage (DAS) devices. It is designed for the corporate and enterprise market as a replacement for parallel SCSI, allowing for much higher speed data transfers than previously available, and is backward-compatible with SATA drives. (SATA drives may be connected to SAS controllers. However, SAS drives may not be connected to SATA controllers.) Though SAS uses serial communication instead of the parallel method found in traditional SCSI devices, it still uses SCSI commands for interacting with SAS end devices. Due to the lower loops speed of 3 Gb/s, SAS drives are currently supported only on the CLARiiON AX series arrays.

EMC CLARiiON AX4-5 400 GB 10k and 15k rpm SAS disk drives

SAS is the next generation of the Small Computer System Interface (SCSI). Parallel SCSI has been the standard server and workstation internal disk storage interface for over 20 years, and was also the standard device interconnect for open systems external array storage prior to the advent of native Fibre Channel drives.

The market most likely to see early SAS adoption is the entry external storage market, which has traditionally utilized either parallel SCSI or SATA drive technology. For these systems, the ability to mix plug-compatible SAS and SATA drives is a distinct advantage over parallel SCSI, which lacks tiered storage flexibility. This plug compatibility (tunneling SATA protocol over SAS) is an area in which SAS technology is still maturing; as a result storage system products are currently entering the market with SAS support only.

The performance difference between 10k and 15k rpm SAS disk drives is similar to the performance difference between 10k and 15k rpm Fibre Channel drives described previously. Details about other capacity SAS drives supported on CLARiiON systems can be found in the *EMC AX Series Storage Systems Disk and FLARE OE Matrix* and the *EMC AX4-5 Series Storage Systems Disk and FLARE OE Matrix* on EMC Powerlink.

Serial ATA hard drives

ATA has traditionally been used for internal storage interconnect in desktop computers to connect the host systems to hard drives and optical drives. Today, the ATA interconnect technology has evolved for much higher interconnect speeds, scalability, and reliability, surpassing the technology's originally intended applications. ATA technologies are now extensively used in enterprise class storage and server environments in near-line storage applications where scale and costs are primary selection driving criteria. Serial ATA (SATA) II is the next-generation internal storage interconnect, designed to replace earlier ATA technologies (SATA I). This interconnect technology is capable of communicating at speeds of 300 MB/s and is the technology of choice used in CLARiiON storage systems.

EMC CLARiiON 1 TB 7,200 rpm SATA II drives

Now we will look at some of the applications where we can implement 1 TB 7,200 rpm SATA II drives. Use SATA II drives in applications that best suit higher-capacity environments to realize true cost and capacity benefits. Sequential access applications have a tendency to maximize any caching advantages of the storage system, and take advantage of the higher density SATA II disk drives. In these applications, the speed at which the drive can transfer data from the platter has the greatest effect on overall performance. Thus, medium-to-large block and sequential I/O applications reap the greatest benefits from the drive's higher area-bit densities.

Some popular types of applications for the 1 TB 7,200 rpm SATA II drives are:

- **Disk-to-disk backup**
  - Disk backup using traditional backup software. CLARiiON with ATA is tested and supported with most major backup applications.
  - Improves backup and restore performance when compared to tape.
- **Large application datasets**
  - Some applications, like seismic data interpretation, government intelligence, and life sciences research are immediately written out to tape due to their large size. When the tests need to be rerun, the data must be reloaded from tape, and then rerun. Now the information can stay online with CLARiiON ATA drives and businesses can improve their operational efficiency and time to market.

Details about other capacity SATA II drives supported on CLARiiON systems can be found in the *CX4 Series Storage Systems Disk and FLARE OE Matrix*.

Benefits of 1 TB SATA II disk drive technology

- Lower price per megabyte
- Reduced footprint
- Reduced power and cooling requirements

EMC CLARiiON 1 TB 5,400 rpm SATA II drives (lower power)

EMC introduced a lower spindle speed 1 TB SATA II drive that delivers the highest density at the lowest cost and requires 96 percent less energy per terabyte than the 15k rpm FC drives and 32 percent less than traditional 7,200 rpm SATA II drives. This drive has been designed to maximize power savings for large, high-capacity deployments such as array-based backup to disk, online tape replacement, and data warehousing. It is ideal for applications where low power usage and cost-to-performance ratio is important and high performance is not a priority.

The drives are available as 15-drive bundles pre-populated in a CX4-4PDAE for new systems, or as upgrades to CX3 and CX4 systems. Installation in the same enclosure with other Fibre Channel or SATA drives model is not supported.

Performance difference between 7,200 rpm and 5,400 rpm 1 TB SATA II drives will depend on the actual application as well as the system configuration, including the number of drives and the RAID type deployed. In highly sequential, large-block environments such as EMC Disk Library, performance of the 5,400 rpm drives has been shown to be comparable to the 7,200 rpm versions of 1 TB SATA II drives.

Detailed analysis of performance characteristics and differences between 1 TB 5,400 rpm SATA II drives and 1 TB 7,200 rpm SATA II drives can be found in the *EMC CLARiiON Performance and Availability: Release 28.5 Firmware Update – Applied Best Practices* white paper available on EMC Powerlink.

Implementing 1 TB SATA II disk drives

EMC recommends that SATA drives be used for single threaded, large block streaming applications. In a typical binding operation, various RAID groups and LUNs are bound and presented to the host as logical disk drives. Within this binding/assignment operation, the hard disk drives are selected and grouped into usable storage for host applications.

It is common practice to mix drive types and enclosures in CLARiiON storage systems, according to user requirements. This is where the 1 TB drives may be factored into the data capacity/performance mix. With the current capacity points of the CLARiiON disk drives at 73/400 GB EFD, 146/300/450 GB (15k rpm), 400 GB (10k rpm), and now the 1 TB 5,400/7,200 rpm SATA II disk modules, you can apply these different capacity and performance drives/enclosures to suit the various applications within your operating environment.

SATA II "Northstar" disk drive and enclosure technology

The EMC UltraPoint™ disk drive enclosure is the current generation CLARiiON disk-array enclosure (DAE) that replaces legacy 2 Gb/s FC loop technology DAEs that were used in the CLARiiON 2 Gb/s CX300/CX500/CX700 series product line. The current UltraPoint design supports Fibre Channel Arbitrated Loop (FC-AL) interconnect speeds of 4 Gb/s between the storage system and/or other DAEs. The FC loop within the DAE is implemented with point-to-point technology. This technology:

- Emulates a loop for FC control traffic (loop primitives, initialization, and so forth).
- Provides a point-to-point connection to each drive for data traffic. This feature provides improved isolation of data traffic error conditions while reducing the loop latency of the FC data.

The UltraPoint storage solution consists of the following field-replaceable units:

- 3U rack-mountable DAE (UltraPoint enclosure) with a midplane and a 15-disk drive module capacity
- Hot-swappable 1-inch low-profile FC disk modules
- Hot-swappable link controller cards (LCCs)
- Hot-swappable blower/power supplies

The SATA II disk drives in this enclosure incorporate a SATA II to FC bridge "paddle card" attached to the back of the EMC CLARiiON SATA II disk drives. This bridge card resides on the disk drive module assembly, and combined with a SATA II drive, emulates a single FC drive.

The main benefit of this technology is that you no longer need to separate “ATA” type enclosures within the CLARiiON storage, because the new “Northstar” SATA II disk drives can be incorporated with standard UltraPoint-style Fibre Channel disk drive enclosures (DAE4P). This allows the CLARiiON back-end loops to continue to run at 4 Gb/s.

Note: SATA II Northstar disk drives cannot be mixed with standard EFD/FC disk drives in the same enclosure.

The implementation of a CLARiiON ATA enclosure is transparent to CLARiiON software. Note, however, that there are some limitations when implementing these ATA enclosures into the CLARiiON CX series systems. The limitations are:

- RAID groups bound on ATA drives cannot span outside ATA drive enclosures, but can span through ATA enclosures. In other words, ATA drives cannot be bound with FC drives in a RAID group.
- Hot spares for ATA drives must be located in ATA enclosures and cannot be used as spares for FC disk modules. (We recommend one spare ATA drive for every 30 ATA drives.)
- ATA drives cannot be located within the first enclosure of the CX series storage system.

- ATA drives should not be used for host booting activities due to performance concerns.

Other than these limitations, ATA drives may be used in any configuration.

When planning for performance and capacity in an environment, the performance, pricing, and capacity requirements for each application should be completely understood before selecting the drive(s). Select the 1 TB SATA II drives when they meet both the storage capacity and performance requirements the individual applications demand.

Disk drive performance comparisons

The following points should be noted about different drive types:

- *Rotational speed* — The SATA II drives spin at 5,400 or 7,200 rpm, whereas the FC/SAS drives spin at 10,000 or 15,000 rpm. EFDs do not have any moving parts.
- EFDs do not have any moving parts; therefore they do not have rotational and seek latencies.
- *Average seek times* — The averages are quite a bit slower on the SATA II drives when compared to the FC drives.

When you take into account the bits per inch (BPI) and internal transfer rates of the two drives, several performance characteristics of the 1 TB SATA II drives become apparent. With average read and write seek rates of 9 ms to 10 ms, and an average rotational latency specification of 4.1 ms, the *CLARiiON SATA II drives may not be well suited for random I/O environments,* such as database or OLTP environments. Instead, the CLARiiON SATA II drives are ideal for bringing offline information online. Offline applications use large, sequential-type data access and storage activities. Some of these applications are discussed later in this paper.

Competitive advantages of the CLARiiON with ATA

Following are some of the competitive advantages to implementing CLARiiON SATA II drives into a new or existing CLARiiON infrastructure. A CLARiiON SATA II drive:

- Offers the full software functionality of CLARiiON with SATA II technology
- Can nondisruptively expand capacity by adding SATA II technology to an existing CLARiiON CX, CX3, or CX4 series storage system
- Provides dual porting on each SATA II drive for high availability
- Makes hot plug and hot swaps available for all CLARiiON ATA components
- Includes FLARE advantages, such as Data Integrity Checking and Scrubbing (Sniffer)
- Provides redundant high-availability components (for example, power, cooling, LCCs)
- Includes the same RAID support (0, 1, 1/0, 3, 5, and 6) and drive-selection flexibility as Fibre Channel drives
- Uses the CLARiiON Sector Data Protection Scheme
- Provides a checksum architecture for end-to-end data protection

Implementing various technology CLARiiON disk drives

In a typical binding operation, various RAID groups and LUNs are bound and presented to the host as logical disk drives. It is within this binding/assignment operation that the hard disk drives are selected and grouped into usable storage for host applications.

It is common practice in the CLARiiON storage systems to mix drive types according to end-user requirements. With the current capacity points of the CLARiiON disk drives at 73/400 GB (EFD), 146/300/450 GB (15k rpm), 400 GB (10k rpm), and 1 TB (7,200 rpm and 5,400 rpm), you can apply these different capacity and performance drives to suit the various applications within your operating environment.

A few examples of mixed disk drive usage

Following are examples of mixed disk drive usage:

- In a CAD/CAM environment, the logical choice of a hard disk drive is higher-capacity storage per drive—for drawing retention and design change—as opposed to lower-capacity storage per spindle.
- In a price-sensitive environment, the lower-capacity hard disk drives may be the right choice due to the lower initial cost of ownership of storage in these types of applications.
- In high I/O, small block applications such as OLTP, EFDs or 15k rpm drives are a good fit from a performance standpoint.

If we take the CLARiiON modular design into account, we could then accommodate all four scenarios into one CLARiiON storage system by simply installing:

- A DAE or RAID group with 1 TB, 7,200 rpm drives for our near-line disk backup-based applications.
- A DAE or RAID group with 1 TB, 5,400 rpm drives for our CAD/CAM environment for drawing retention.
- A DAE or RAID group with 400 GB, 10k rpm drives for our engineering-group applications.
- A DAE or RAID group of high-performance 73/400 GB EFDs and 146 or 300 GB, 15k rpm drives for OLPT and random database applications.
- A DAE or RAID group with 400 GB 10k rpm disk drives for a good midpoint-cost/performance-effective storage solution.

CLARiiON disk drive power solutions

CLARiiON's advanced storage technologies improve overall system performance and optimize storage energy use. In the following sections we discuss the foundations of energy usage and CLARiiON best practices for improving energy efficiency.

Multi-tiering

CLARiiON has the unique ability to make effective use of multiple tiers of storage capacities within the same system. CLARiiON's UltraScale architecture allows a wide variety of disk drive technologies within the same array. A single UltraPoint disk-array enclosure (DAE) can support multiple drives (either SATA or EFD, FC, and/or Low-Cost Fibre Channel) and multiple interface speeds (2 Gb/s and/or 4 Gb/s). The flexibility and configuration options provided by use of UltraPoint technology, in conjunction with Virtual LUNs, enable you to easily move between any of these drive technologies.

It is common practice in the CLARiiON storage systems to mix drive types according to end-user requirements. This is where 1 TB 5,400/7,200 rpm SATA II drives should be factored into the data capacity/performance mix. With the current capacity points of the CLARiiON hard disk drives at 73/400 GB (EFD), 146/300/450 GB (15k rpm), 400 GB (10k rpm) and 1 TB (5,400/7,200 rpm), you can select the best drive to suit the capacity and performance requirements of each application in your environment. Figure 3 shows the power savings with different capacity versions of FC and SATA II drives.

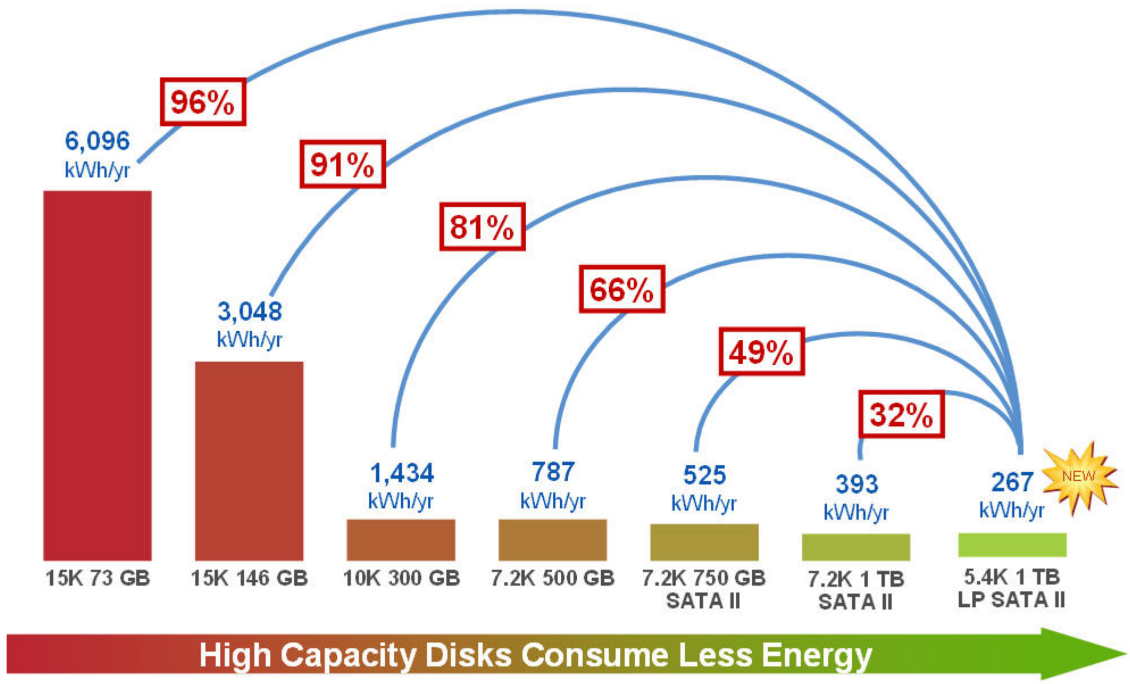

Figure 3. Power savings with different capacity FC and SATA II drives for 1 TB raw capacity

The ability to move data dynamically from one tier to another in the CLARiiON storage system results in significant energy savings as different storage tiers have a different power profiles. Table 2 shows the power consumption profile of different types of disks that are available in CLARiiON. Note that EFDs have very little difference between active and idle power consumption because of the absence of any moving parts in the drive.

Table 2. Power usage between active and idle drives

Rotation speed	Active power	Idle power	Difference between idle and active power
15k rpm	24.15 W	20.99 W	13%
10k rpm	23.35 W	18.91 W	19%
7.2k rpm	21.35 W	11.51 W	46%
5.4k rpm	16.98 W	10.57 W	38%
EFD	15.02 W	15.02 W	0%

It is crucial to consider several factors when planning for performance, capacity, and energy requirements for any given environment. Select the proper disk drive drives that meet the storage capacity, price point, power consumption, and performance requirements that your individual applications demand. Table 3 lists the power consumption for different disk drive technologies in a standard CLARiiON environment. Note that these figures compare only the drive technology choice and not their relative energy per capacity.

Table 3. Typical power consumption of different drive technologies

Drive type	Number of drives	Line current	Power consumption	Annual energy costs	Heat dissipation
15k rpm	15	1.83 A	0.381 kVA	$976	1,240 Btu/Hr
10k rom	15	1.77 A	0.368 kVA	$943	1,200 Btu/Hr
7.2k rpm	15	1.62 A	0.337 kVA	$863	1,100 Btu/Hr
5.4k rpm	15	1.29 A	0.268 kVA	$686	870 Btu/Hr
EFD	15	1.14 A	0.237 kVA	$607	770 Btu/Hr
Enclosure power consumption calculations based on a local utility rate of 0.1537 $ /kW-hr					

Using one disk drive type for all load types on a storage system is not a recommended practice, although some vendors may propose this practice. Each storage tier has a different information and power requirement throughout its lifecycle. Thus, the different tiers of storage should be a major consideration when deciding on system configuration/layout for storing and managing data.

Rebuild times

Rebuild times depend on a number of factors, including the rebuild rate setting (ASAP, High, Medium, or Low), presence/location of an appropriate hot spare, disk type, disk size, bus speed, application load, and RAID group topology. When a hot spare is used for the rebuild, there is an additional “equalize” operation that occurs when the faulty drive is replaced and the content from the hot spare is copied to it.

Equalize rates on RAID 5 are faster than the rebuild itself since only two drives are active (the hot spare and the new drive) and there is no need to recompute parity or data from the other drives. Equalize on RAID 1/0 is identical to the rebuild, only in the opposite direction. The effect of ASAP rebuild on application degradation depends on the workload mix. If an ASAP rebuild cannot be tolerated by an application, rebuild time can be paced with the High, Medium, or Low setting. These rebuild rates are much slower than ASAP, but at the High setting production workloads are competing with the rebuild only for 10 percent of the time.

Baseline ASAP rebuild/equalize rates for common RAID group configurations using 15k rpm FC drives on a single 4 Gb/s loop can be found in the *EMC CLARiiON Performance and Availability: Release 28.5 Firmware Update – Applied Best Practices* white paper on EMC Powerlink. Information about how different priority settings affect the array’s performance during a rebuild process is also available in this paper.

Performance planning

Performance planning or forecasting is a science that takes considerable knowledge. The steps presented here are intended for rough estimation only.

Rule-of-thumb approach

To begin performance estimation, a rule of thumb is used for IOPS per disk drive and MB/s per disk drive. This is a conservative and intentionally simplistic measure. It should be noted that this is only the beginning of an accurate performance estimate; estimates based on the rule of thumb are for quickly sizing a design. More accurate methods are available to EMC personnel.

The approach for a quick estimate is:

- Determine host IOPS or bandwidth load.
- Calculate disk IOPS or bandwidth load.
- Calculate the number of disk drives required for disk IOPS or bandwidth load.
- Calculate the number and type of storage systems.

Rule-of-thumb numbers for all the CLARiiON drive technologies and speeds are explained in *EMC CLARiiON Performance and Availability: Release 28.5 Firmware Update – Applied Best Practices* on EMC Powerlink.

EMC Hard Drive Reliability Qualification

In general, EFD, FC, and SAS hard drives have "Enterprise" level reliability. SATA drives have near-line reliability. Prior to releasing any new technology, such as EFD and SAS, several technology compatibility and characterization tests are conducted. In all cases, the drives are subjected to extreme environmental stresses and a great deal of performance parameters are collected.

Figure 4 shows a summary of the steps EMC takes to move a disk drive from definition to manufacturing and a customer readiness phase for general availability.

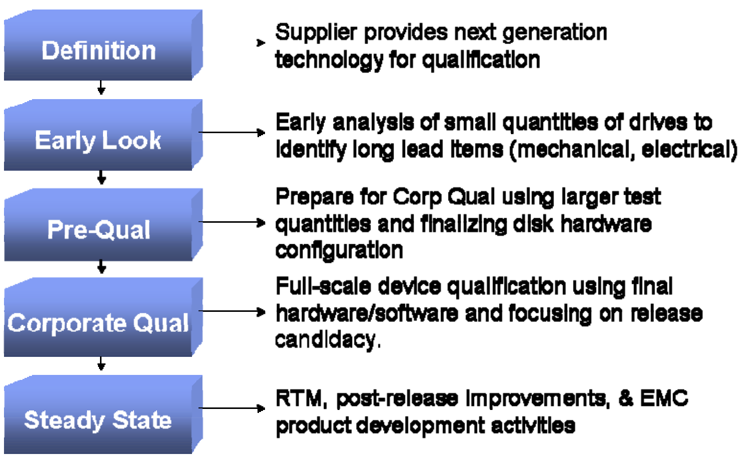

Figure 4. Taking a disk drive from definition to general availability

EMC has established a world-class disk qualification and reliability process to ensure adherence to the highest standards for our customers.

At the conclusion of testing, if the results are satisfactory and are approved by the EMC management teams, the drive then ships in volume under the guidelines of EMC's manufacturing process controls.

EMC takes deep pride in the level and degree of testing before a disk drive is released for manufacturing. Only after going through these exhaustive lists of processes and qualification processes is a drive then sent into manufacturing.

Conclusion

EMC provides several different disk drive technologies for various applications, capacities, cost points, and use cases. In all cases, EMC performs extensive qualification and testing to ensure longevity and reliability required by enterprise applications.

As a result of the varying capacities and performance characteristics currently supported on the CLARiiON storage systems, we are now able to support four types of deployments—meeting four different price/performance tradeoffs within one CLARiiON storage system platform.

We provide our customers the highest standards in overall system performance, scalability, and reliability.