



Enterprise-class SATA

The new alternative for enterprise storage applications

Why SATA anyway... a brief history

The driving force behind the definition and launch of SATA can be summed up by two disparate factors.

First was the need for speed as highest IDE/ATA disk transfer rates were fast approaching the maximum 100MB/sec parallel-bus specification and would reach the IDE bus theoretical performance limit of 133MB/sec soon thereafter. To address the performance limitation and ensure a long run in the market the SATA specification defined a high speed serial interface roadmap of 1.5 Gigabits (Gb) per second initially, followed by 3Gb/sec and then 6Gb/sec. Due to an 80% efficiency encoding from parallel IDE/ATA this translated to 150, 300 and 600MB/sec respectively. Compounding the approaching performance limit was that desktop disk drive capacity and transfer rates (areal density + RPM) would soon push the disk data transfer rate beyond the IDE/ATA bus limit. This speed mismatch historically was a big no-no for disk drive vendors because of risk of inconsistent performance if the disk drive buffer overflowed and a revolution was skipped. As it turned out the performance loss was negligible which in turn delayed the acceptance of slightly higher priced SATA disks within the desktop market.

The second and more market-driven factor behind the definition of SATA was to enable low-cost IDE/ATA drives to be sold into enterprise applications for use in secondary storage applications and as tape replacement for backup. For those disk drive vendors with no SCSI/FC product offerings or enterprise customers, SATA would enable the possibility to sell into new markets with higher profit potential than the razor-thin margins within the desktop market. Therefore the SATA specification also defined a number of interface enhancements, specific for use in high-end enterprise applications, some of which were originally pioneered by SCSI disk drives. These include:

- Blind-mate or hot-plug connectors to allow for active sparing in RAID environments
- Native command queuing (NCQ) with data scatter/gathering to improve random I/O performance
- 32-bit cyclic redundancy checking (CRC) of errors for both data and control commands
- Improved cabling for increased robustness and overall system reliability
- Point-to-point topology to eliminate potential arbitration delays and provide higher reliability

It is important to note that until SATA disks were adopted into select enterprise applications, the overall volumes remained small due to slightly higher component costs and no compelling value proposition for typical desktop applications. Driving storage costs down in select enterprise applications was the market catalyst for the initial SATA volume not improving desktop performance.

The introduction of enterprise-class SATA disk drives

One common misconception about disk drives is that many consider the interface and the disk drive to be one and the same for determining reliability and performance. But it is important to understand what mechanisms have the greatest impact. Both the interface (i.e. SCSI, SATA, etc.) and what can be called the disk drive engine (i.e. head-disk assembly (HDA), RPM speed and average seek times) contribute to overall performance but the engine; specifically the HDA, is the predominant factor in determining reliability, especially after the SATA interface enhancements listed above were implemented.

But the SATA specification addressed only the interface performance and reliability, not the disk drive engine, which remains largely vendor specific. This has allowed disk vendors to independently advance their SATA HDA designs and deploy what can be considered enterprise-class SATA disk drives.

Some of these enterprise-class advancements now designed into SATA disk drives are:

- SATA disk drives are now designed and tested for use in high duty-cycle environments
- Larger data caches (8MB now typical) for improved performance
- Fluid dynamic bearings (FDB) for improved long-term reliability, reduced heat, vibration and noise
- Time limits on error recovery to prevent timeout conditions with some RAID controllers
- Increased mean-time-to-failure (MTTF)¹ specifications to between 1 million to 1.2 million hours

Breaking down the myths & misconceptions

Reliability:

One myth is that there is a tradeoff in reliability when adopting SATA disks and the reason they are lower in price as compared to SCSI disks is that they are not built to the standards of SCSI drives. Disk suppliers with significant SCSI volume have vested interest in prolonging this misconception within the enterprise market as long as possible to preserve their higher margin products revenues. However, for the most part this reliability perception is simply not accurate. It is true that earlier generations of desktop-class IDE/ATA disks were rated for about 600,000 MTTF but today's enterprise-class SATA and SCSI disks are typically over 1 million hours. Therefore, it is reasonable to expect the reliability of SATA hard drives to approximate that of SCSI drives. Most SATA disks are also designed and tested for continuous duty in RAID arrays. When enterprise-class SATA disks are installed in a RAID configuration with redundancy, cooling, and other high-reliability features designed into the system, there is statistically no difference in the mean-time-to-data-loss (MTTDL) between SATA and SCSI disks, with RAID-5 MTTDL predictions for either interface in the tens of thousands of years.

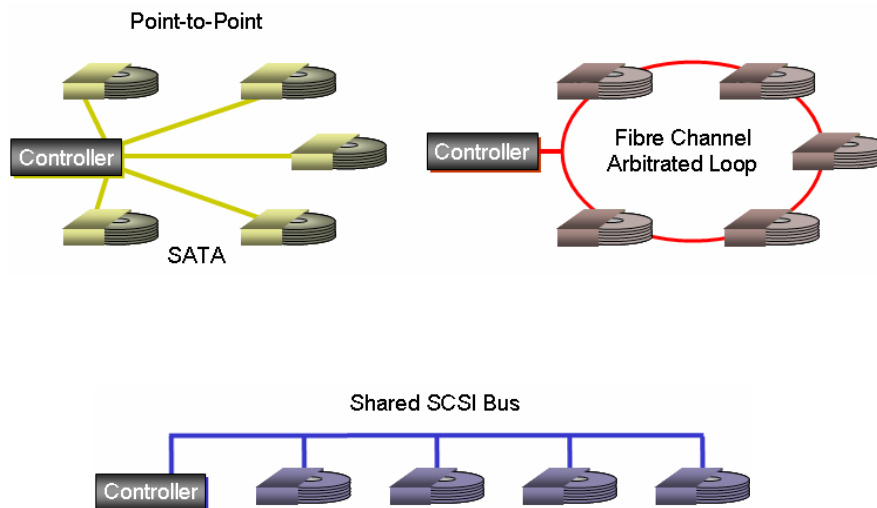
Disk Vendor	Model	Disk Interface	MTTF (AFR)	Error Rate	Operating Shock	Operating Vibration
Western Digital, Caviar RE	WD1600SD	SATA	1M hrs	1 in 10 ¹⁴	65G	0.75G
Western Digital, Caviar RE	WD2500SD	SATA	1M hrs	1 in 10 ¹⁴	65G	0.75G
Maxtor Maxline Plus II	7Y250M0	SATA	>1M hrs	1 in 10 ¹⁵	60G	1.0G
Western Digital, Raptor	WD740GD	SATA	1.2M hrs	1 in 10 ¹⁵	65G	3G
Seagate Barracuda	ST3250823	SATA	.34 AFR	1 in 10 ¹⁴	63G	0.25G
Maxtor Atlas IV	8B146J0	SCSI	.73 AFR	1 in 10 ¹⁵	63G	1.5G
Hitachi UltraStar	10K300	SCSI	1.2M hrs	1 in 10 ¹⁵	15G	1.0G
Fujitsu	MAT3300	SCSI	1.2M hrs	1 in 10 ¹⁵	65G	1.0G
Seagate Cheetah 10K.6	ST3146807	SCSI	1.2M hrs	1 in 10 ¹⁵	25G	.5G

This table illustrates the similarities and narrow differences between both SATA and SCSI disk specifications².

Other metrics in determining reliability and most commonly used by the disk drive vendors are annualized return rate (ARR) and annualized failure rate (AFR). As indicated in the table, some disk vendors only display reliability in terms of AFR. The disk AFR specification is the preferred method for communicating reliability specifications to volume OEM customers and ARR provides a mechanism for determining how specific disk drive models are performing in the field. Disk vendors' datasheets often specify particular models using either AFR or MTTF which tends to complicate the true reliability comparisons between SATA, SCSI & FC products. But in general .73 AFR equates to approximately 1.2M hours MTTF for enterprise-class disks³. Using these methods, recent data from select disk vendors indicate the AFR of enterprise-class SATA disks are performing at lower .43% as compared to the .73% SCSI ARR specification.

Performance:

Many disk vendors position SCSI superior in performance to SATA, but the opposite can be true. Sustained data transfer rates, which are important in streaming high-throughput environments, are in the range of 60-70MB per second for today's SATA disks and compare favorably to all but the fastest, most expensive SCSI disks. But because of its lower command overhead and point-to-point topology, where every SATA disk is directly attached to the controller, SATA disks can actually out perform SCSI disks in streaming mode. Also, there are no bus contentions or hang issues in a SATA point-to-point architecture. Both SCSI and FC-AL (Fibre Channel Arbitrated Loop) are shared bus architectures, where each disk must wait its turn to send or receive data and creating potential throughput or even access problems for other disks on the same bus. SATA point-to-point architecture enables even small numbers of disks to provide hundreds or even thousands of aggregate MB/sec of throughput without bus contention. In on-line transaction processing (OLTP) applications where high RPM and fast seek times are required, effective striping of data across multiple SATA disks and controllers provides very high levels of performance.



Gauging the future for enterprise-class SATA

First, the future is serial. The trend toward serial technology for all server data transport mechanisms is unavoidable and becoming pervasive. Beyond FC there is PCI Express along with Infiniband and now SATA. Even SCSI is evolving from a parallel bus to serial with Serial Attached SCSI (SAS). There is no escaping the synchronization constraints of parallel data buses with high-speed serial links the only viable data transport platform moving forward. This trend should indicate the end is coming for parallel SCSI with some disk vendors no longer offering SCSI on future products. Although SAS and 2.5-inch remain distant wild cards, the two dominant enterprise disks architectures are likely to be 15K RPM FC for extremely high end performance applications and cost-effective SATA (both 7,200 & 10K RPM) for almost all other applications.

Enterprise-class SATA disks combine the benefits of high volume commodity-driven desktop IDE/ATA disk drive component pricing, but with enterprise-focused design for reliability, performance, hot swappability, adaptability to RAID arrays and manufacturing process. Additional enterprise functions and requirements should be expected in the future to further enhance SATA disks. As a result, not only are companies deploying enterprise-class SATA in high-throughput secondary storage applications such, as disk-to-disk (D2D) replication for business continuity and disaster recovery and as virtual tape libraries, but also in primary storage applications within high-end NAS, where SATA coupled with effective RAID striping provide high levels of operations per second performance and enterprise-class reliability.

In the 20 years since RAID was created, the "inexpensive" in the acronym evolved to "independent," reflecting the evolution toward expensive ultra high RPM disks to achieve higher RAID performance. SATA point-to-point topology deployed in systems with multiple RAID-5 disk groups and hot sparing enables solutions with enterprise reliability, performance and pricing that puts the "inexpensive" back in RAID.

As with any disruptive technology specific disk vendors, depending on their SCSI/FC business and cannibalization concerns will be more or less aggressive in designing and promoting SATA disk drives with enterprise reliability and performance attributes. Disk vendors with no current SCSI/FC business or market presence are expected to take the lead to exploit the compelling price/performance advantages of SATA and expand an ever growing acceptance of SATA disks in enterprise storage markets and applications.

¹ Alternately referred to as mean-time-between-failure (MTBF) and provides information about a population not a particular drive.

² Information obtained from the respective disk vendors web sites and datasheet specifications

³ Sourced from Seagate Cheetah 10K.6 drive specification

Western Digital and Caviar are registered trademarks of Western Digital Technologies Inc.
 Maxtor, Maxline and Atlas are trademarks and registered trademarks of Maxtor Corporation
 Seagate, Barracuda and Cheetah are registered trademarks of Seagate Technologies LLC
 Hitachi and UltraStar are registered trademarks of Hitachi Global Storage Technologies
 Fujitsu MAT is a registered trademark of Fujitsu Computer Products America