Oh Data, Where are Thou?
The Case for ATA and SATA Reliability

Thomas M. Coughlin
Coughlin Associates
www.tomcoughlin.com

The Ecclesiastical Consumer

In vain we toil but our productions seem destined to enter the valley of darkness. The destructive forces of this world will eventually turn everything we buy into mud, dust and ash. But long before that we will have already bulldozed them into landfills because it costs too much to fix a 5 cent power switch. Fortunately, we can rest assured that long after our BMWs are crushed, our Barolo bottles smashed and our houses flattened by space junk, that photographs of us in our most embarrassing moments will still be available for viewing somewhere on the Internet – because even if all our stuff gets chewed up and scattered, the Internet is the ultimate, reversible information spittoon.

Lest we wander too far off track, prognosticating the silent end of hollow men buried in the depths of a towering wasteland, at least we can take some satisfaction in our ability to make complex products like automobiles that can be made to last several decades if we can only find a mechanic with intuition and a timing light. The question in our consumptive "*product is the FRU*" economic world is how much does a product cost and how much 5-year risk is assumed when buying it? And by the way, where have all those electric typewriters gone anyway? Maybe they are buried beneath that pile of unmatched and discarded socks in that scary bedroom drawer.

Seeing as how we are all storage bozos on this bus, this article takes us on a drive down the backroads of Disk Drive country. Our mission: understanding drive failures better so we sleep at night with our money saving ATA and SATA based storage subsystems.

Why, Oh Why Do Disk Drives Fail?

Hard disk drives (HDDs) are sophisticated electromechanical devices that integrate many diverse technologies including precision motors, electronic circuitry, magnetic media platters and advanced read/write heads. With so many constituent components involved, there are many potential ways a disk drive could fail.

Figure 1 presents the general layout of a hard disk drive showing the rotating spindle and voice coil (head actuator) motors, the head suspensions, disk media platters, castings and cover, as well as the drive circuit board.

Disk drives have various failure mechanisms that lead to their demise, as catgorized in **Table 1**. Nearly all of these failure mechanisms result from the operation of the disk drive, but some of them can also be caused by innate material weaknesses, aging or the manufacturing process. Often, more than one failure mechanism can be in operation in a drive family at one time.

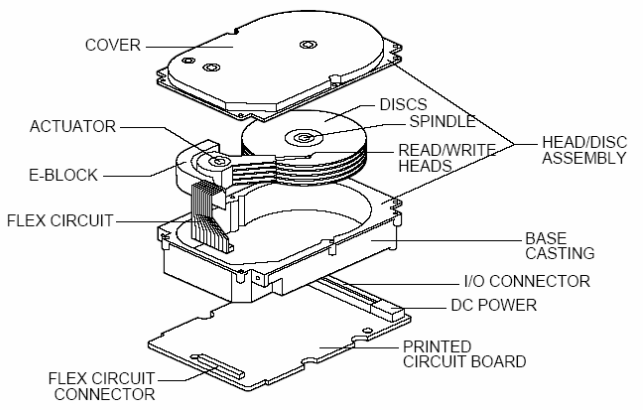

Figure 1. General Layout of Hard Disk Drive Components

Table 1. Disk Drive Failure Mechanisms Indicating Operating or Non-Operating Failures

General Failure Pathology	Failing Component	Cause of Failure	Failure Origination
Corrosion	Media Head	Contamination of media Contamination of heads	Operations/Materialse Operating/Materials
Media Errors	Head Media Head	Thermal asperity-head hits bump on disk Embedded defect in media from manufacturing High-fly writes due to lubricant accumulation on the head	Operations Materials Operations
	Head/Media Media	Loose hard particle damage to heads or media Head slaps that create defects on media	Operations Operations
Head Instability or Dead Head	Head Head	Repeated contact with thermal asperities Head asymmetry from manufacturing	Operations Materials
	Head Head Head	Many long writes Hard particle contact Electrostatic Discharge	Operations Operations Operations/Materials
Hardware Errors	Motors Bearings Mechanics Other Mechanics	Various Various Rotational Vibration Various	Operations Operations Operations/Materials Operating
Electrical Failures	Printed Circuit Board DRAM ICs Motor Drivers Pre-Amp	Solder or connector failure Various Various Various Various	Operations/ Operations Operations Operations Operations

Living in a Bathtub

The classical way to represent the time dependent failure rate of a complicated system is to assume that there are three different failure modes that occur at different times during the average lifecycle of a disk drive, as shown in **Figure 2**. Sometimes this analysis is referred to as the "bathtub curve" These three ranges of the bathtub curve are:

1) Early life (also known as infant motality), failures, which result primarily from marginal materials or manufacturing anomalies and typically occur within hours of initial operation.
2) Service life failures which result neither from materials/manufacturing defects or from accumulated wear and tear, but represent other random failures, and;
3) Wear out failures which result from expected wear and tear of system components.

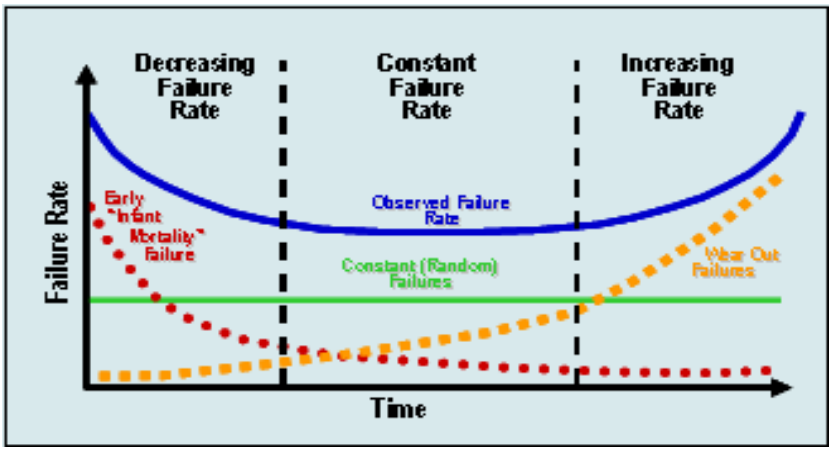

Figure 2. Failure Rate Bathtub Curve

In *the early life or infant mortality region* the disk drive experiences early life failures or "infant mortality." These failures are characterized by components that pass the product specifications and testing process but fail shortly after voltage or other normal operating stresses are applied to the drive. Early life failures occur usually within the first few hours of operation.

In *the service life region* the marginal components have already failed and most units can be expected to operate successfully for several years as long as they are used per the specified operating conditions and environment. Disk drives are designed so that almost all of the disk drives will be operational after five years.

In *the wear out region* the disk drive has usually experienced several years of successful service but normal wear and tear begin to cause components in the drive to operate outside of their specifications. Although this wear and tear usually occurs slowly at first it gradually accelerates and within a short period of time the failure of one part leads to the failure of other parts and an avalanche of failures occurs that ultimately leads to total disk drive failure. As mentioned before for modern disk drives used within their specificed operating conditions and environment region 3 failures usually don't occur for at least five years after the disk drive has begun operation.

Disk Dryve Disk Dryve, Burning Brite...

Hard disk drives generally undergo what is called a "burn-in" as one of the final operations in manufacturing a hard disk drive. During burn-in the completed disk is powered up and placed in a chamber that may be at a somewhat elevated temperature in order to induce early failures. For several hours the drive is exercised with the disk spinning and the heads seeking across the disk surface writing and reading information. The purpose of the burn-in is to weed out drives that may be susceptible to early life failures and prevent them from causing problems for customers.

Can't Take the Heat

With most infant mortality problems being caught by storage vendors, drives reaching the field should fail primarily due to service life failures (region 2) , which are not very frequent. Failures in this phase of the lifecycle are generally activated and/or accelerated due to environmental factors although, as will be pointed out later, usage can play a very important role in disk drive expected life. Probably the largest single accelerating factor for disk drive failures is the ambient temperature around the disk drive and the related internal disk drive temperature.

If the ambient temperature is raised above room temperature the rate of failure of a hard disk drive increases. These failure rates are often expressed as annual failure rates or AFR. An acceleration factor is applied to the basic failure rate equation to estimate the increase in drive failure rate as the temperature in raised. **Figure** 3 is an example of the increase in failure rates (acceleration of failures) as a function of temperatures above normal room temperature (25 C)[1].

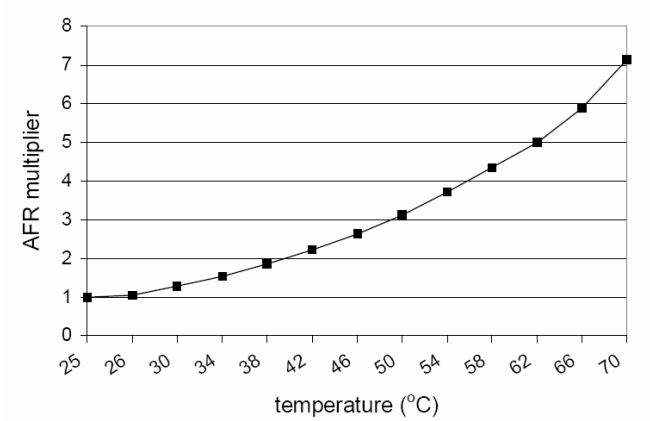

Figure 3. Reliability Decrease Due to Ambient Temperature Increase[1]

Power by the Hour

Other usage factors influencing disk drive failure rates are similarly analyzed. **Figure 4** is an example of annual failure rate acceleration due to the annual power on hours[1]. This is generally interpreted as the failure rate accelerator dependent on the assumed number of hours that the hard drive is powered up and operating.

---

[1] **Estimating Drive Reliability in Desktop Computers and Consumer Electronics Systems**, G. Cole, Technology Paper TP-338.1, Seagate Technology, November 2000.

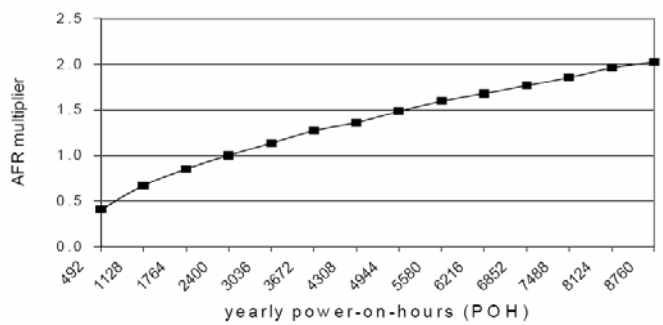

Figure 4: Reliability Decrease as a Function of Annual Power-On Hours[1]

Drive usage is a big factor in drive failure. The simple act of regularly turning a disk drive on and off during normal operations can significantly shorten the resulting life of the hard disk drive. Power cycling can stress electronics and spindle motors, leading to early failures. Note that some disk drive power saving modes that turn the disk spindle motor off while keeping the drive electronic powered up avoid some of the stress on the circuit boards of the drive. Some modern hard disk drive storage systems such as Massive Arrays of Inactive Disks (MAIDs) achieve significant power saving and drive longevity by balancing drive power saving modes with expected data retrieval characteristics for an archived file.

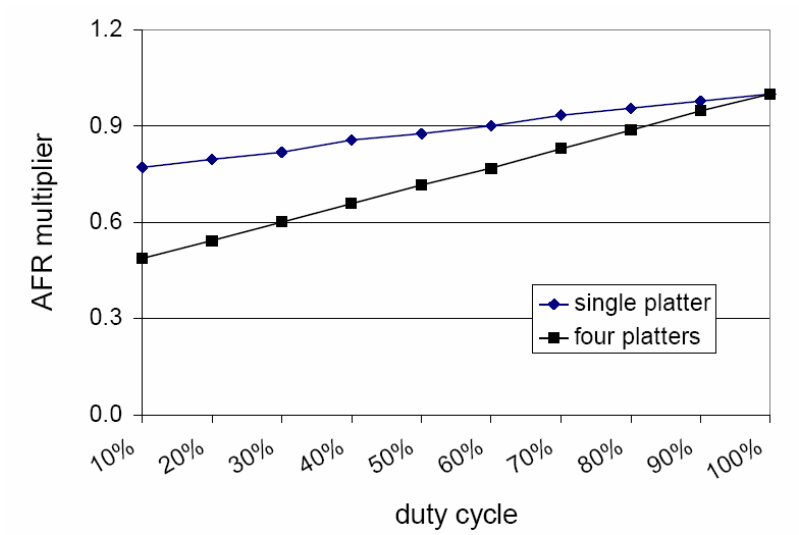

Figure 5. Reliability Decrease as a Function of Duty Cycle for One and Four Platter Disk Drives

Another usage factor is a measure of the amount of mechanical work that the disk drive does, otherwise known as the duty cycle. **Figure 5** is an example of the annual failure rate multiplication factor as a function of the percentage of the time that the drive is actively writing and reading information[1]. As shown in this figure the failure multiplier goes up as the numbers of platters (and heads) increases. The failure rate decreases as the duty cycle percentage decreases.

Bad Actors Competing for Attention

Alas, things are not as simple as they seem at first glance. Often there are competing failure mechanisms at work in a family of hard disk drives. These different failure mechanisms may each behave very different with time. In fact there often are very pronounced changes in the failure rate of disk drives as a function of time, technology, and production vintage.

Figure 6 shows the observed failure or hazard rate for a disk drive family that was found to have at least three separate failure mechanisms (FM1, FM2-FM5, and FM6) acting together to create the composite failure rate[2]. FM1 appears to be early failure rate due to infant mortality that after a peak early in life declines with time. FM2 through FM5 appear to be the same, a steady-state low level failure rate. FM6 looks like a wear-out failure rate that increases with time. The resulting overall drive failure rate is also plotted showing complex behavior where the failure rate initially increases with time, flattens out for a while before rising steadily again as time goes on.

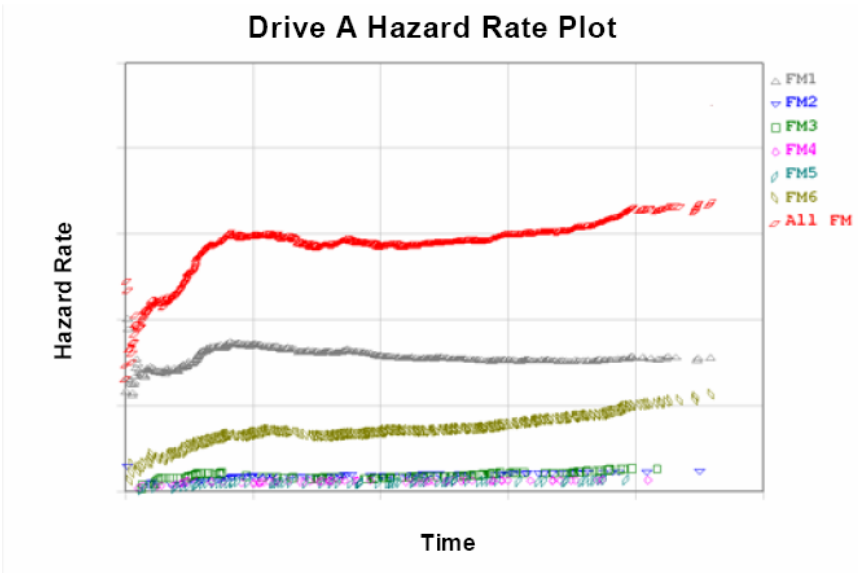

Figure 6 Hazard Rate Plot of Failure Mechanisms for a Drive Family[2]

In the case of the drives represented by Figure 6, FM1 are failures caused by handling or contamination problems which tend to cause early life drive failures and decrease as the disk drive manufacturing processes mature and as the drives are longer in the field. FM2-FM5 are random failures, which dominate during the useful (or service) life of the disk drive. Failure mechanism FM6 is likely due to early wear-out of one or more components in the drive due to marginal specifications for the components or to a drive design or operation that causes some components to wear out prematurely.

The Telescoping Law of Diminishing Failures

There are three different perspectives on failure rates: the consumer's, the integrator's (such as a computer or subsystem manufacturer) and the manufacturer's. Disk drive

---

[2] **Reliability Analysis of Disk Drive Failure Mechanisms**, S. Shah & J. G. Elerath, Proc. Annual Reliability & Maintainability Symp., Jan. 2005, pp 226-231.

companies get most of their returns from system and storage subsystem manufacturers. Historically, SCSI and Fibre Channel drives installed in high end systems and storage subsystems have been more likely to be returned for credit than low cost ATA and SATA desktop hard disk drive installed in desktop systems. Drives that are not returned are not counted by the drive manufacturer as a failed drive.

In addition, drive companies may choose to not count drives as failures if they believe other problems or abuses may have occurred. For instance a disk drive that shows signs of damage due to mishandling is usually dismissed as having died from unnatural causes. Drive manufacturers categorize mishandled and abused drives as "non-chargeable failures" or NCFs. In addition, a significant percentage of returned drives are tested and found to be working well. These "no trouble found" or NTF drives sometimes appear in the market as refurbished product.

The result of combining the NTF rate with the NCF rates results in failure rates measured by the drive company that are approximately 35% of those observed by the end customer. As a result of these accounting differences, the drive company recognizes an MTBF rate that is 2.78 times higher than that observed by the end customer.

Tell me Doctor, what Condition is my Condition in?

These plots of failure rates shouldn't scare customers unduly. General disk drive reliability has increased significantly over the 50-year life of the hard disk drive industry and continues to improve.

Certainly the reliability and error recovery capabilities of individual drives is important, but the design of a storage subsystem can change the reliability equation significantly. The use of mirroring or parity RAID significantly increases the reliability of storage well past the reliability expectations of a single drive. In addition, there are other techniques that can be applied to reduce the likelihood of experiencing fatal disk errors. For instance many subsystems employ background *disk scrubbing* techniques that constantly look for read errors on disks in the subsystem. If an error is encountered, data can be rewritten or relocated to other disk locations that are known to be good. Standard technology incorporated in most drives today called SMART (*Self-Monitoring Analysis and Reporting Technology*) is also used to spot degrading conditions in disk drives. An intelligent subsystem controller can respond to SMART information and take corrective action, such as employing subsystem spare disks to make additional copies of data. These things can occur transparently as a subsystem background task without requiring any action on the part of the storage administrator.

In Yo Face Wid Da Interface

There has been a long standing debate about the reliability of ATA and SATA disk drive used in RAID-based disk drive arrays vs. SCSI or Fibre Channel (FC) high performance disk drives that have traditionally been used in enterprise storage applications. As shown in previous sections, the lifespan of a disk drive depends on many factors such as design, environment and usage conditions.

SCSI and FC disk drives offer faster time to data on the drives due to higher rotational rates (up to 15,000 RPM), faster seek time to data tracks due to more powerful motors, and smaller disk platters. They also cost a lot more than ATA and SATA drives. Conventional wisdom says that if the highest random access performance (which means a high duty cycle) is required for an application and price isn't a constraint, SCSI or FC disk subsystems are the optimal choice.

If more reading than writing is required, as is the case for fixed-content data, then an ATA or SATA based storage subsystem is almost a slam dunk decision. One of the big trends in storage today is to use ATA or SATA subsystems as secondary storage in a tiered storage environment. Data stored on them is readily available and the difference in access times compared with SCSI and FC drives are mostly indiscernible in therse applications. ATA/SATA SAN and NAS subsystems are also popular choices for SMB businesses that cannot justify the higher costs of SCSI or FC storage. As more SATA subsystems are deployed, customers are becoming more comfortable with them and less wary of the risks. Also, as SATA technology evolves, the designs are going to include more features related to their use in storage subsystems. Most SATA vendors have already started incorporating these assumptions into their products and the results are likely to be excellent for customers.

There are many opinions on whether SCSI and FC disk drives are more reliable than SATA drives designed for enterprise storage environments. David Anderson, et. al[3] from Segate have written that SCSI and FC disk drives use certain components that are more costly, but provide higher reliability. He also points out that different firmware and electronics on these drives result in a richer command set that supports the incorporation of additional reliability features. The claim is also made that SCSI and FC drives are more thoroughly tested before going to market and that this additional testing is done to decrease service life failures as well as early life failures.

At the end of the day, the basic question that people want an answer to is this: "How good is good enough for my applications?" For starters, no disk drive is designed to fail. On the contrary, they are designed to be very reliable because disk drive vendors cannot afford to ruin their reputations with problematic products. Disk drive vendors have manufactured both server and desktop drives for many years that are designed to keep their customers satisfied and maintain business relationships in a brutally competitive industry.

Manufacturers have always used different assumptions about the applications for server (SCSI/FC) and desktop (ATA/SATA) disk drives. The Seagate team's assertion that first generation SATA drives were not designed to be used in server environments is true. However, it needs to be said that a storage subsystem can be a much better environment than either a server or desktop system. Cooling, power, vibration, controllers and connectors can all be vastly improved in a subsystem design. The majority of vendors making SATA subsystems have been extremely sensitive to the reliability concerns of the market and have worked hard to make reliability a non-issue for their products. Not only that, but SATA drives in a rack mounted subsystem running in a cooled server room are

---

[3] **More than an interface—SCSI vs. ATA**, D. Anderson, J. Dykes, E. Riedel, Proceedings of the 2nd Annual Conference on File and Storage Technology (FAST), March 2003.

in a much friendlier environment than those same drives would be if they are in a desktop system stuck under the desk of a toe tapping, knee knocking office worker. Finally, the duty cycle of a drive in a subsystem may be much less than if that same drive were running inside a server system, depending on the design of the subsystem.

Hughes and Murray[4] state that since ATA/SATA disk drives have an unrecoverable error rate specification that is an order of magnitude worse than that of SCSI or FC drives, and since they generally have a greater storage capacity, drive reconstruction failure rates are worse for the ATA/SATA RAID subsystems. They argue that double parity RAID (N+2) should be used with SATA subsystems in order to increase reliability so as to match or exceed that of SCSI or FC subsystems.

In practice, subsystem failures of SATA subsystems have been rare, although the conclusions of Hughes and Murray's paper predicted very high failure rates. Its not that SATA drives have not failed – they do occasionally just as SCSI and FC drives sometimes fail – but the occurrence of multiple drives failing in an array appears to be quite low. Something is obviously going on to make the average SATA disk drive much more reliable than many people feared. And that thing is good engineering by the makers of SATA subsystems (For more on this topic see the sidebar: "Assumptions, Damn Assumptions and Reliability Statistics).

Desktop disk drives are usually specified at about 8 hours a day of operation since that is the typical work day and the computers containing the disk drives are usually turned on in the morning and off at the end of the day. Therefore people wrongfully believe that leaving drives these running on a 24X7 basis is hard on them. But this assumption is not necessarily valid because the simple act of turning drives on and off stresses the drives and itself can be a cause of premature failure. A SATA drive in a disk subsystem that never shuts down does not have the problem of periodic power surges coursing through its integrated circuits every day. Considering that IC failures are approximately 8 times greater than spindle failures, it follows that leaving them running all the time actually improves their reliability.

SCSI and FC disk drives are specified for 24 x 7 operations and have MTBFs in excess of 1,200,000 hours. Desktop ATA disk drives are specified with a MTBF of about 600,000 hours Desktop ATA and SATA drives have a 1-3 year warranty while enterprise SCSI and FC drives have a 5 year warranty.

In principle, there is no reason why an ATA or SATA drive should have lower reliability than a SCSI or Fibre Channel disk drive if used in the right environment, with the right firmware, factory testing and optimization, in a properly designed subsystem, and under the appropriate usage conditions. Desktop ATA and SATA disk drives can be used in enterprise environments where spare drives are available to accommodate possible early wear out of the drives.

Disk drives should be chosen based on the desired system performance and budget. There is room for several types of storage components in the enterprise environment.

---

[4] **Reliability and Security of RAID Storage Systems and D2D Archives Using SATA Disk Drives**, G. F. Hughes & J. F. Murray, ACM Trans. On Storage, Vol. 1, No. 1, December 2004, pp. 95-107.

Tiered storage using a combination of high performance SCSI drive subsystems combined with higher capacity, lower priced, and somewhat lower performance SATA subsystems (at least for random writes) can be a reliable solution to an organization's general storage needs.

Final Byte (Summary and Conclusions)

Disk drives are complex systems and the analysis of their reliability requires considerable knowledge and control or compensation of various failure mechanisms. These failure mechanisms are quite numerous but they generally fall into three categories: early life failures, random failures during the drive service life, and end of life wear-out failure. During the service life of the disk drive the failure rate tends to be low and relatively stable. In this operating region failures are often represented by an approximately constant failure rate and disk drives and disk based storage systems are often tested to meet a mean time between failure (MTBF) specifications.

In practice a given disk drive family is likely to have several mechanisms of failure that results in a hazard or failure rate curve as a function of time that follows a complex behavior depending upon which failure mechanism dominates at a particular point in time. This complicates the analysis of drive reliability. Over the 50 year history of the disk drive industry drives have been growing ever more reliable under ever more demanding enterprise environments. Given the proper drive and storage system design as well as the proper operating environment and usage conditions, reliable storage systems can be built using SCSI, Fibre Channel, ATA or SATA interface disk drives. Each type of storage device offers trade-offs of performance, capacity, and price and together they provide the basis of tiered storage systems that work together to provide reliable and cost effective enterprise storage solutions.