ASHRAE TC9.9

Data Center Storage Equipment – Thermal Guidelines, Issues, and Best Practices

Whitepaper prepared by ASHRAE Technical Committee (TC) 9.9
Mission Critical Facilities, Data Centers, Technology Spaces, and Electronic Equipment

Table of Contents

Executive Summary

This paper defines thermal guidelines best practices for storage equipment. The intended audience is both storage hardware designers and storage equipment users. The thermal guidelines were written by a team of storage subject matter experts from many different companies and the recommendations are industry-wide best practices which are agnostic of the point of view of any one company.

A comprehensive review of environmental specifications and air flow management best practices was conducted across disk, flash, and tape based storage equipment. The review uncovered disconnects between the environmental specifications of the storage equipment and the building blocks inside it such as hard disk drives, super-caps, and batteries. In many cases, the building blocks weren't capable of meeting the published system specifications, especially for non-operational conditions. Mis-alignment of sub-assembly and system level specifications could result in a customer experiencing IT equipment failures and data loss even though they were following the published environmental specification for their storage equipment. ASHRAE proposes a set of industry standard recommendations for disk, flash, and tape to align system and sub-system environmental specifications and prevent possible equipment failures and data loss.

Cooling air flow management practices, both inside the storage equipment and at the rack and data center level, were investigated. One simple finding was, for data center design and operation, customers need equipment documentation to provide a range of air flow for different levels of equipment power levels, not just a maximum air flow. Data centers designed based on maximum air flow numbers will be over-designed and less energy efficient. Customers and data center operators focused on energy costs need detailed tables and graphs for airflow versus system loading.

Another opportunity for equipment energy savings is the implementation of more sophisticated fan control algorithms. Recent generation servers and storage arrays have step based fan control algorithms or even polynomial fan control based on internal temperature sensor feedback. Sophisticated sensor-based fan control algorithms can do a better job of "right sizing" cooling air flow to maximize energy savings. At a data center level, ASHRAE recommends the adoption of multiple environmental control zones as a best practice. Where possible, equipment with the same or similar environmental control requirements should be grouped together in the same zone or zones. A zone based approach to cooling and physically locating equipment will realize power and cooling cost savings for a majority of the data center while making sure critical equipment, such as tape or batteries, is controlled within narrower specified environmental limits.

Other important findings are the need for periodic or real-time monitoring of gaseous pollutant levels. Sources of gaseous pollution can be seasonal and variable. Periodic monitoring of gaseous pollution is needed to determine in advance when copper and silver corrosion rates are reaching a level that could cause equipment failures over time.

Data center sites should be chosen, where possible, to minimize external vibration, such as construction and railroad traffic. High levels of transmitted vibration have been shown to impact hard disk drive performance. Design of data center infrastructure should minimize internally generated vibration from sources such as movement of heavy equipment, air movers, and loud noise from fire alarm horns and fire suppression systems.

1 Introduction

This paper is written for a broad audience that includes data center power and cooling experts as well as IT and storage equipment specialists. Sections of the paper may be basic for some audience members. The reason for this is to provide a wide base of background information that bridges any gaps in understanding and provides a common framework for understanding the proposed storage thermal guidelines, issues, and best practices.

One projection for the rate of information storage growth is given below in **Figure 1**. Another plot that shows the breakout of storage growth by media type is shown in **Figure 2** Although the rate of storage growth varies slightly from one reference to the next, all of the sources show the need for information storage is growing almost exponentially **[1] [2] [3] [4]**. According to IDC, the world's information is doubling every two years. By 2020 the world will generate 50 times the amount of information we generated in 2011.

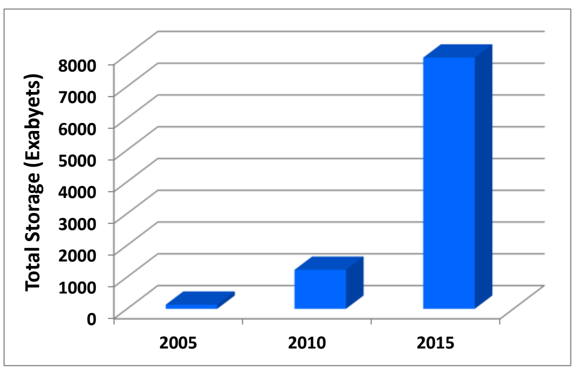

Figure 1 Growth of total world-wide storage from IDC's Digital Universe Study (2011) [3].

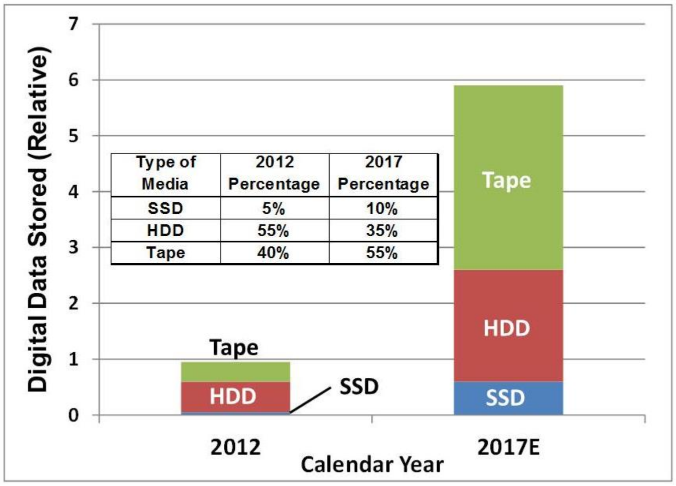

Figure 2 Breakout of storage growth by storage media type [1].

The amount of data generated, collected, processed, and stored by most organizations will continue to grow aggressively for the foreseeable future and this will drive rapid growth in information storage hardware.

Growth in data storage is being fueled by three broad categories [5]: a) social media, b) machine data, and c) transactional data. The growth of social media data is being fueled by the large volume of images and videos that are generated from mobile devices. Machines are generating more data than ever before. For example, the Large Hadron Collider (LHC) generates 40 terabytes of data every second during experiments [5]. Even a simple transaction with an on-line retailer can generate data a significant amount of data including product IDs, prices, payment information, manufacturer and distributor data, and customer behavior.

The information storage trend in **Figure 1** may be affected by software data compression technologies that would reduce the amount of storage volume required for the same data. Software data compression technologies already exist and are already in widespread use [6], [7]. Any change to the projected growth of information storage would be brought about by either the adoption of new software based data compression methods or by wider adoption of existing data compression technologies. Even with data compression, the expectation is still for rapid growth in the amount of data stored.

1.1 Primary Functions of Information Storage

In a data center there are four primary functions of information storage: on-line storage, backup, archiving, and disaster recovery.

Table 1 Summary of the four primary functions of information storage.

Storage Function	Description

On-Line	Rapid access, frequently used, current information
Backup	Guard against media failure and data loss
Archiving	Long term storage of data needed only infrequently
Disaster Recovery	Protects against loss of information from natural disaster such as tornado. Disaster recovery storage is typically done at a separate physical site

1.2 Types of Storage Media and Storage Devices

Information storage is currently done by means of several different types of media including hard disk drive, solid state disk drive, tape and optical media. Some of these storage media are shown in the figures below.

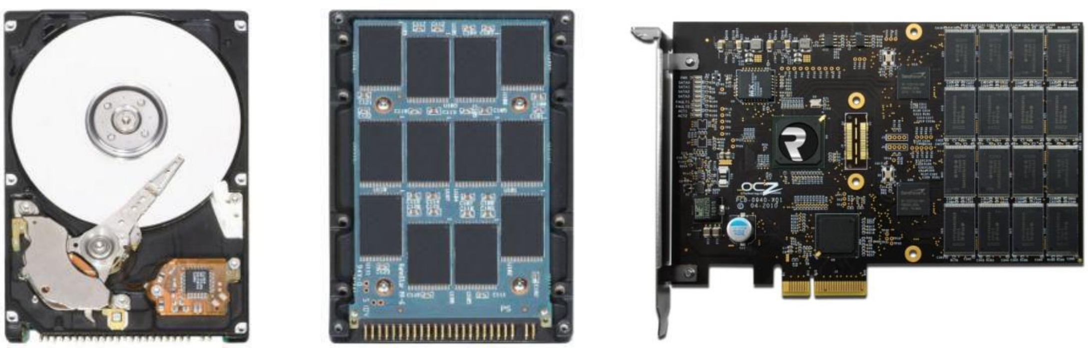

Figure 3 Hard disk drive with cover removed (left) compared to a flash memory based solid state drive (center) and a Peripheral Component Interconnect Express (PCIe) based solid state drive (right).

A Peripheral Component Interconnect Express (PCIe) SSD shown above in **Figure 3** plugs into a PCIe slot instead of a drive bay. PCIe allows for much faster data transmission rates than the connection to a typical drive in a drive bay.

Figure 4 Example of a single tape drive (left) and a tape media cartridge (right).

Solid state drives use non-volatile flash memory for information storage whereas traditional hard drives use platters of rotating magnetic media. A comparison of active power across different types of HDD and SSDs is given below.

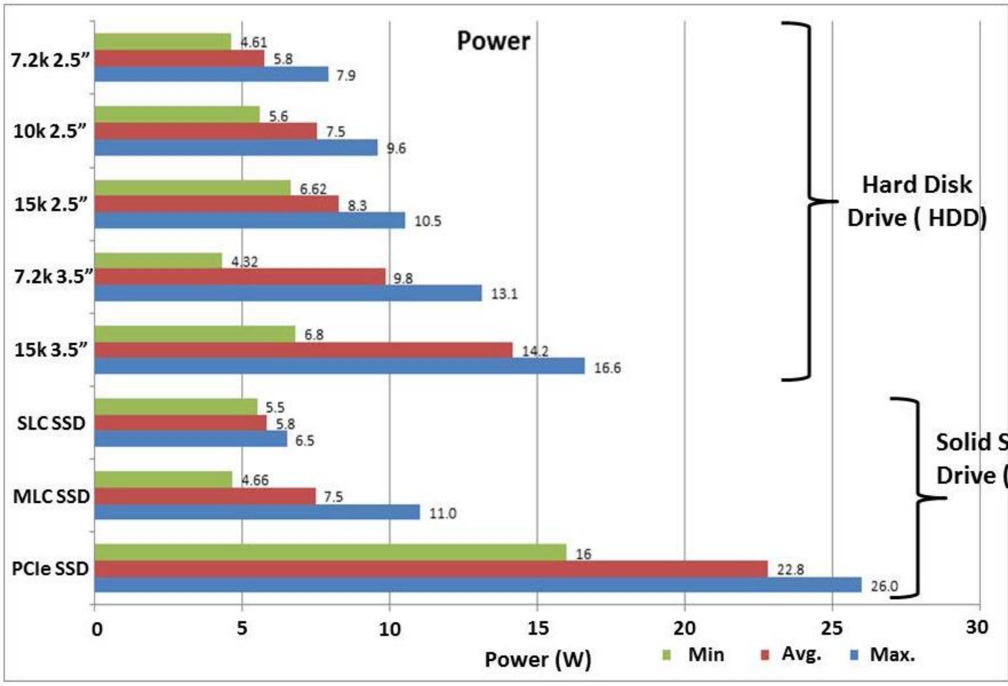

Figure 5 Comparison of the active power consumption of various sizes and types of drives.

Among hard drives, 2.5" drives have an average power consumption of around 6W per drive with slightly higher power consumption of 7.5 and 8.3W at higher spindle speeds of 10K and 15K, respectively. 3.5" drives have an average power consumption of approximately 10 watts for a 7.2K spindle speed and a power consumption of over 14W per drive for a 15k spindle speed. The power consumption of solid state drives, whether single-level cell (SLC) or multi-level cell (MLC), is very similar to 2.5" hard drives. However, PCIe based drives have a much higher power consumption, ~23W, than any of the other drives.

Power consumption is an important criteria for selecting a type of storage media. Some types of media have unique energy usage models that make direct comparisons between media difficult. For example, tape consumes energy when it is in a drive being actively being written or read but the energy consumption an individual tape cartridge in a tape library is almost negligible.

A quick summary of the positive and negative attributes of commonly available storage media is given below.

Table 2 Summary of the pros and cons of the commonly available storage media.

Type of Storage Media	Pros	Cons
Hard disk drive (HDD)	Fast read and write access to data. Mature technology.	HDD's can suddenly crash. RAID or other strategies are often required to protect against data loss.
Solid state drive (SSD)	Fast read and write times. PCIe bus SSDs are even faster. SSD's degrade slowly over time and generally don't have sudden unrecoverable crashes like HDDs.	Higher cost. Flash memory has a limited write lifetime.
Tape	Most reliable technology for long term data archiving. Low cost, mature technology, minimal energy usage for offline storage	Relatively slow write and read access times
Optical Media	Inexpensive, readily available.	Not suitable for enterprise applications. Disks can be affected by mechanical damage.

Error! Reference source not found. quantifies the performance of each media type in key areas such as energy usage (active and idle), data rates, hard error rates, and shelf life

Type of Storage Media	Enterprise Hard Disk Drive (HDD)	Enterprise 2.5" Solid State Drive (SSD)**	Tape	
			Enterprise	Midrange
Active Energy Usage Range (W)	4 to 17	4 to 11	50 to 90	27
Energy Usage at Idle (W)	0.6 to 11	2.8 to 4.1	30 to 36***	7.5***
Max Sustained Data Rate (MB/sec)	115 to 370	500 to 750	75 to 250	40 to 160
Hard Error Rates (bits)	1 in 10^{16}	1 in 10^{16}	10^{20}	10^{17}
Shelf Life (years)	Not defined	3 months at 40°C	30*	30*

**Assumes tape drive hardware is still available to read the tape media after 30 years.*
***Does not include PCIe solid state drives.*
****These values are for a tape cassette in a tape drive. The energy consumption of a stand-alone tape cassette in a library is almost negligible.*

1.3 Disk Storage Arrays

Disk arrays are categorized based on how they implement redundancy and how they are connected to a host network or computing device.

RAID vs. JBOD

Disk storage arrays are found in two flavors; Redundant Array of Inexpensive Disks (RAID), or Just a Bunch of Disks (JBOD). The latter JBOD is precisely what the name implies; multiple disks in an enclosure that are exposed to a host just as they are. In the case of the former, the RAID device is not what it seems.

A typical enterprise class HDD will have an annualized failure rate (AFR) of between 0.7 and 1.0%, making the HDD one of the most failure-prone components in a storage subsystem. Consequently, steps are taken to ensure that, when an HDD failure occurs, the data stored on the HDDs is still accessible. Data are typically stored across multiple disks. Data are stored and indexed according to a Logical Block Address (LBA) scheme. In today's art, a logical block of data is 512 bytes, and is written to an HDD with a header and error correction code. A hard disk drive can have millions of logical block addresses, as the number of logical block addresses will be approximately equal to the capacity of the drive in bytes divided by the logical block size. The RAID device appears to the host as a physical device, but can actually be a multitude of physical devices. The LBAs of these physical devices are mapped by the RAID controller into a virtual address space that appears to the host as a single device. As data are received by a RAID storage controller to be written to a logical volume, the RAID controller distributes the data blocks across multiple preconfigured physical volumes, using a predetermined stripe size. Striping is the process of writing data across multiple physical disks. When data is written in stripes across multiple physical volumes, a mathematical exclusive OR (XOR) function is used to generate parity data. XOR compares two input bits and generates one output bit. If the bits are the same, the result is 0. If the bits are different, the result is 1. In the case of RAID5, an additional physical HDD is used to store the mathematical XOR parity data of all the data within a stripe. If one of the HDDs in the volume group is lost, the data stored on the failed HDD can be "rebuilt", by again doing a mathematical XOR of all of the remaining data within the stripe.

One of the most common types of data center level information storage is a disk based storage array. The basic hardware components of a disk storage array are shown below in **Figure 6**.

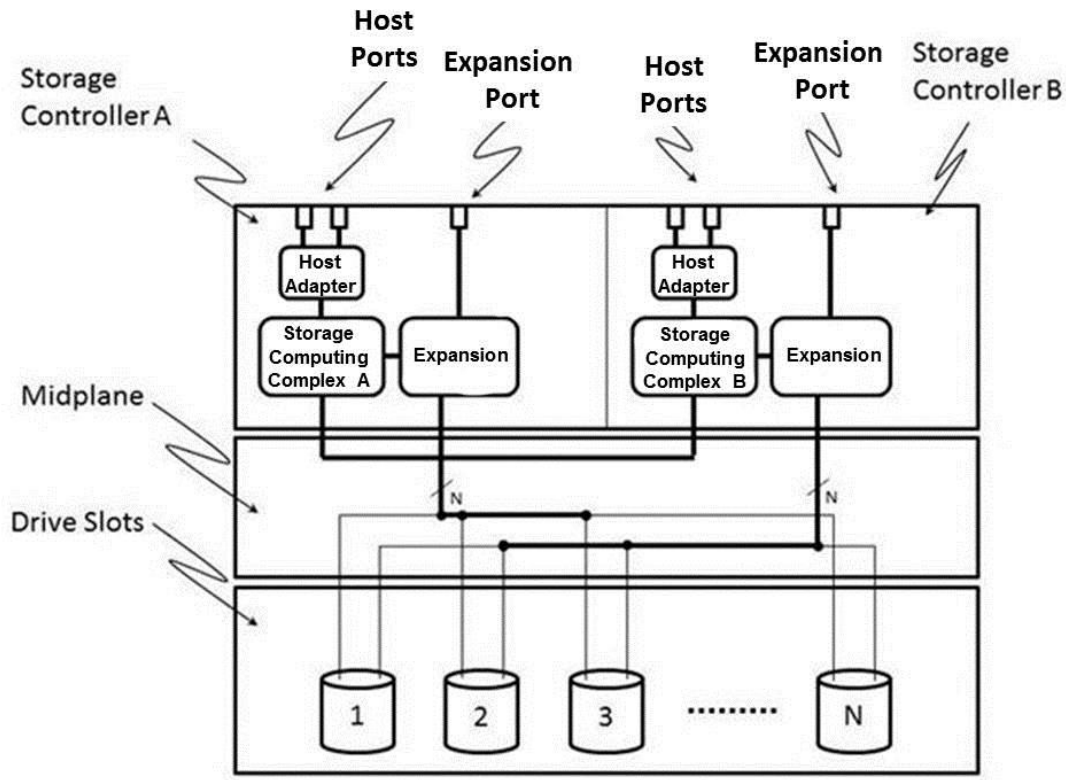

Figure 6 Schematic diagram of typical disk storage array.

The array enclosure comprises two RAID controllers A & B, which provide redundancy for fault tolerance and enhanced connectivity. Each RAID controller contains a Storage Computing Complex, which can be a specialized RAID-On-a-CHIP (ROC) or a general purpose processor with specialized IO controllers (IOC) allowing conversion from standard PCIe to storage centric interfaces such as Serial Attached SCSI (SAS), Serial ATA (SATA), etc. The storage controller is accessed through Host Ports using common storage protocols such as SAS, iSCSI, Fibre Channel (FC), etc. The Storage Computing Complex acts as a block-level command server to SCSI command descriptor blocks issued by a host computer in the course of storing and retrieving data. A Host Adapter converts the external storage protocol to PCIe to be handled by the Storage Computing Complex. The data written by an external host computer is distributed by the Storage Computing Complex to physical media according to the configuration parameters defined by the RAID algorithm and the logical to physical volume mappings. Because the internal storage interfaces are serial point-to-point connections, an expansion module is used to allow the Computing Complex to address multiple physical drives through the enclosure mid-plane; each drive being located in a drive slot and having a dedicated link to the expansion module. The drives each have a primary and secondary port, such that each controller has a connection to each drive. The Storage Computing Complex also has a local cache to temporarily store data for quick read access and more efficient write operations; consequently, there is also a communication path through the mid-plane, allowing the two controller Storage Computing Complexes to share caching information and data as well as Storage Enclosure Service operations.

In addition to the controllers, mid-plane and drives, the Storage Array Enclosure typically employs redundant power supplies and cooling modules in the event of component failures as shown below in **Figure 7**.

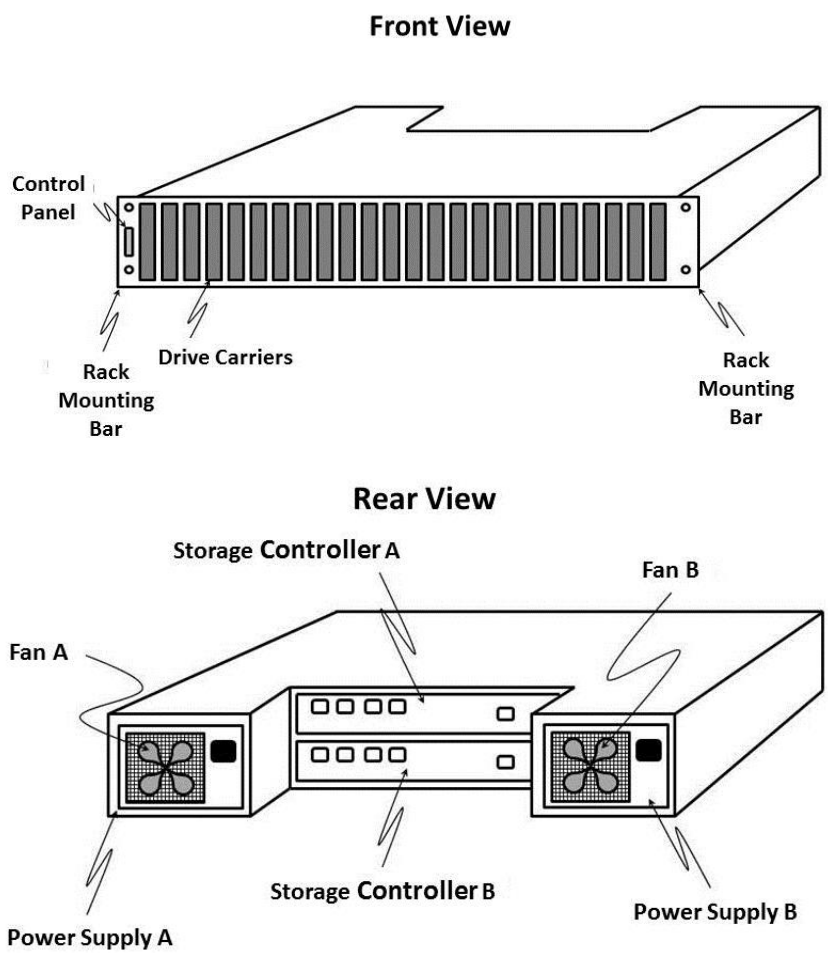

Figure 7 Front and rear views of storage array enclosure.

The deployment and connectivity of storage arrays is typically one of three common types: Network Attached Storage (NAS), Storage Attached Network (SAN), Direct Attached Storage (DAS). The different types of storage arrays are discussed in Appendix B.

Figure 8 below shows how individual disk arrays show in **Figure 6** and **Figure 7** are installed in racks to make a much larger storage array.

Figure 8 Racks populated with 4U height disk storage array modules.

1.4 Tape Storage

Even though tape is a relatively old storage medium it is still very popular and, per **Figure 2**, the amount of tape storage being used is still growing. Tape systems come in a variety of different formats based on the type of tape media.

Linear Tape-Open (LTO) Format

LTO, the abbreviation for Linear Tape-Open, a tape format optimized for high capacity and performance with high reliability. Available in either single or multiple drive configurations, the LTO tape format is ideally suited for enterprise-level backup, restore and archive applications. LTO is the open standard alternative to proprietary tape technologies which results in interoperability across vendors and more economical product line for the user. The latest generation of LTO, Generation 6, holds up to 6.25 TB of data (with 2.5:1 compression). More information can be found on the Ultrium LTO consortium web site [8].

3592 Format

The 3592 format is a series of tape drives and corresponding magnetic tape data storage media formats. They are designed to meet business demands that require applications that rely on high capacity and fast access to data as well as long-term data retention. The 3592 format is used with TS1140 tape drives. A TS1140 drive can store up to 4 TB (native capacity) on a single tape cartridge.

T10000 Format

T10000 is a family of high capacity, high performance, enterprise class tape drives and magnetic tape data storage media formats. T10000 compatible drives are designed for heavy data retrieval and tape library automation. A single T10000 tape cartridge provides for up to 8.5TB (native capacity).

2 Power Density and Cooling Trends

2.1 Power Density Trends

Data center economics tends to favor denser equipment that takes up less floor space. However, higher power density can create challenges for power distribution and cooling. The trend over time for storage and networking IT equipment power density is shown below in
Figure 9. The definition of the heat load per equipment footprint in the graph can be found on page 37 of reference [9].

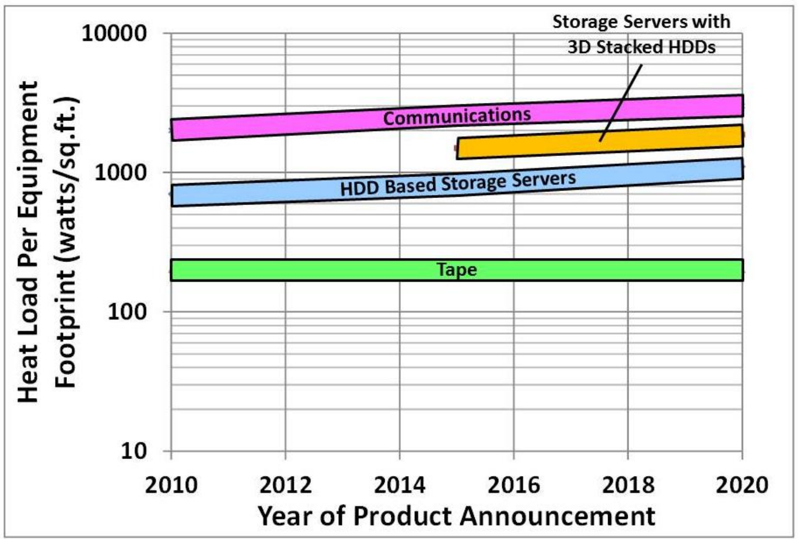

Figure 9 Power density trend over time for storage and networking IT equipment.

The power density of storage array systems is largely determined by: a) the power density of the storage media, b) the amount of storage media used inside an enclosure, and c) the type of storage media being used. The power trend for hard disk drives shown below in **Figure 10** is flat over even slightly decreasing over time, even with increases in hard disk drive density.

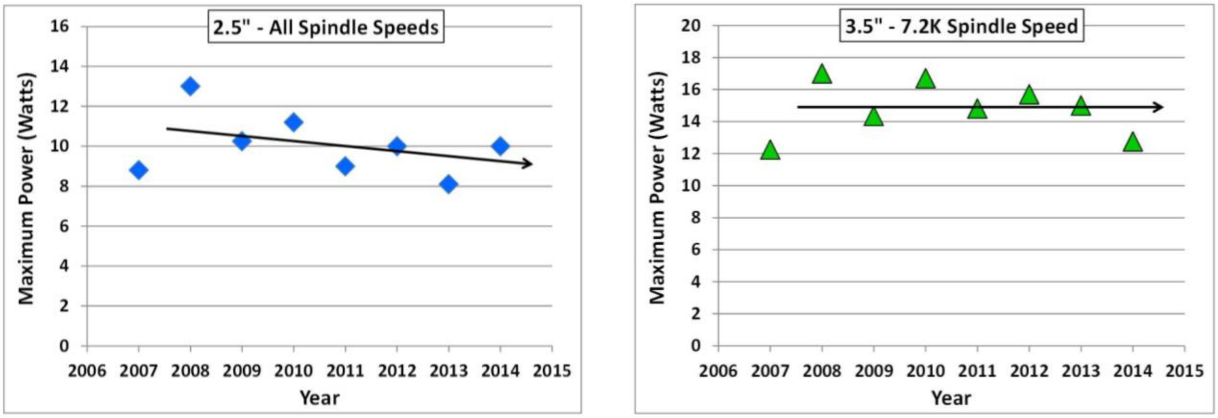

Figure 10 Maximum power trend for 2.5" and 3.5" hard disk drives [10].

However, increases in the power density of storage arrays could still come about if a larger number of drives were designed into an enclosure of the same size. Power density could also increase if, for example, the array was provisioned with high power drives such as 15k spindle speed. Though it is a secondary factor, the choice of RAID levels can also affect power consumption.

Another path where power density increases can come about if the drives are stacked three dimensionally inside in the enclosure. Traditional hard drive and solid state drive based storage arrays place the drives only at the front face of the storage array enclosure for optimal cooling and accessibility. The area of the enclosure face limits the total number of drives. In a three-dimensional storage array, such as the one shown below in **Figure 11** the drives are placed behind one another. Three dimensional stacking of drives can significantly increase the capacity of a storage array but it will also significantly increase the array power density and cooling needs.

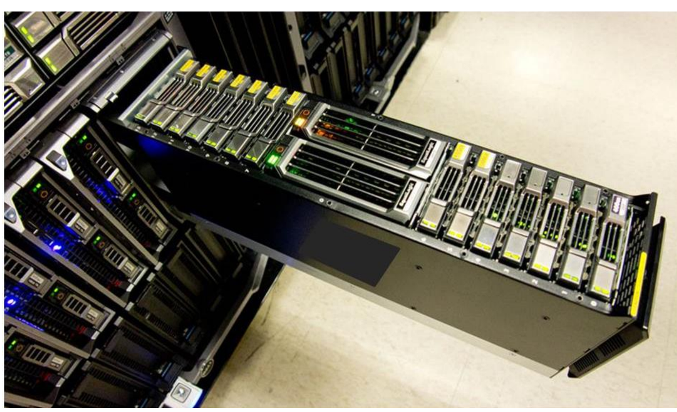

Figure 11 Blade storage drawer with 14 three-dimensionally stacked hard disk drives.

Solid state drives have similar power consumption compared to hard disk drives of similar form factor because they have been designed to fit into existing enclosures with fixed power budgets. PCIe based solid state drives can consume up to 3X more power than a conventional hard disk drive (see **Figure 5**). If a storage array was built using PCIe based solid state drives, the power density of the array could be substantially higher than a conventional storage array.

Some opportunities end users can implement to improve the energy efficiency of storage equipment are:

1. Choose storage equipment that is scalable. Keep just enough storage capacity on-line to meet current needs and keep additional storage installed and ready but in a powered down state until it is needed.
2. Right size the number and types of hard disk drives and solid state drives to minimize power (see **Figure 5**). Use higher power high spin speed hard drives such as 15k spindle speed only when necessary. Where rapid read/write access is needed consider using a solid state drive in lieu of a high spindle speed hard disk drive.
3. Maintain infrequently accessed data ("cold data") in more energy efficient storage such as arrays with low spindle speed drives or even tape.
4. Consider energy efficiency as one of the factors in choosing your RAID array architecture [11], [12].

Overall, the forecast over time is for a storage (tape and drive based storage) power density to increase only slightly over time. However, there may be a few exceptions to this trend such as storage arrays designed with three-dimensionally drives.

2.2 Common Cooling Air Flow Configurations

The primary means of cooling commercially available storage equipment is air. The most common configuration of individual disk drive based storage arrays are the 2U and 4U form factors. In general, disk based storage arrays don't occupy as much rack cabinet depth as a comparable height (2U or 4U) server. However, this depends on whether or not the array has two rows of drives, one behind the other, or three-dimensional packaging. If a storage array has horizontally stacked drives, i.e. 3D, the storage array may occupy nearly the entire depth of the rack. The most common front of rack to back of rack ("front to back") storage array cooling air flow design is shown below in **Figure 12**.

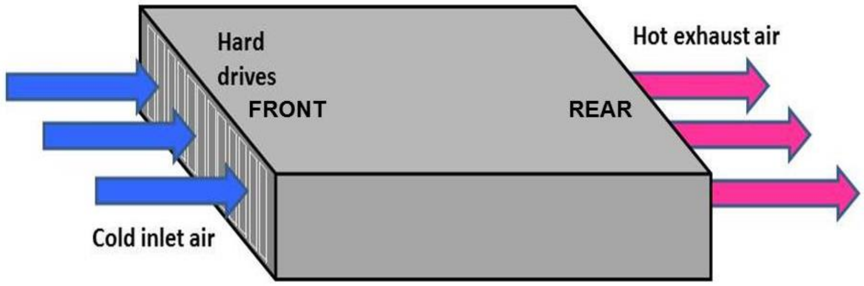

Figure 12 Front of rack to back of rack air flow typical of a 2U or 4U storage array.

Individual 2U and 4U disk drive based storage arrays are often used as building blocks to create full height fully populated rack storage arrays of much larger capacity. The most common rack level cooling architecture consists of front to back air flow as shown below.

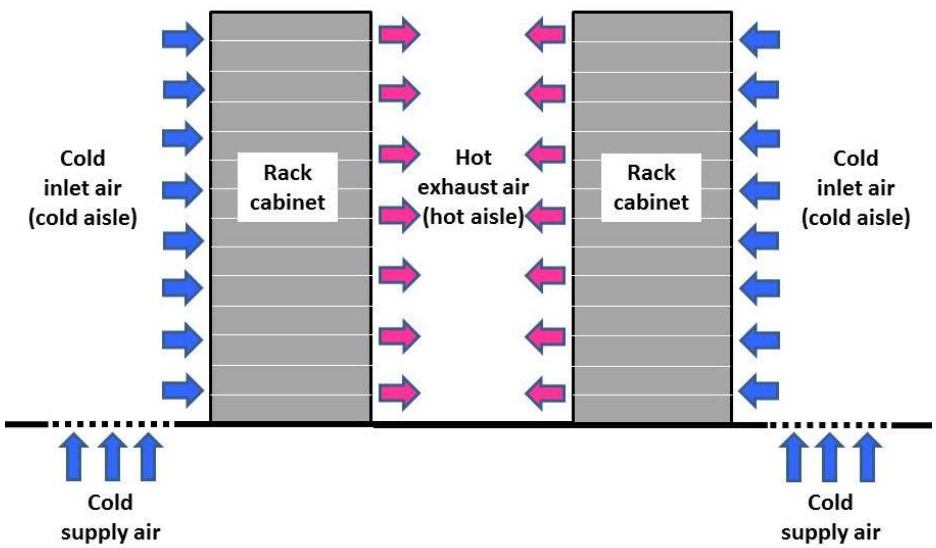

Figure 13 Storage array cabinets with front to back cooling air flow. Hot aisle/cold aisle configuration is shown.

Some storage arrays have a specially designed cabinet door that acts as a plenum to more evenly distribute the supply of cooling air coming through a floor tile to the individual storage arrays along the height of the rack as shown below. Without the plenum, the storage arrays near the bottom of the rack can consume a disproportionate amount of the cooling air leaving the arrays near the top of the rack running hotter.

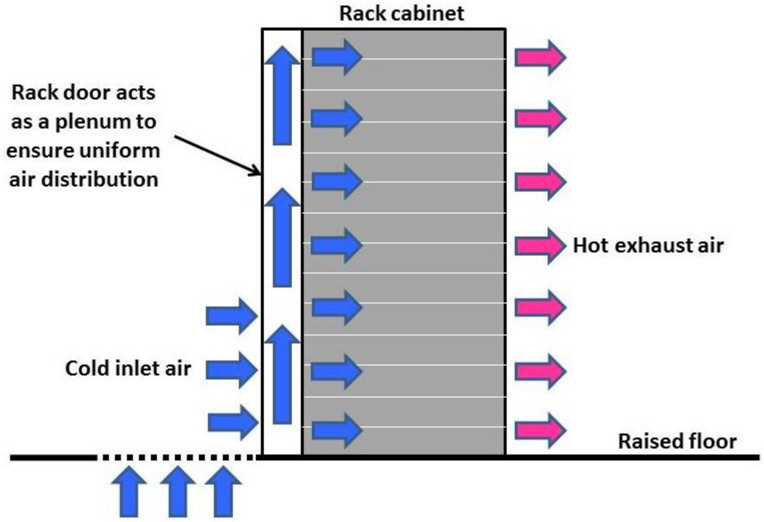

Figure 14 Cabinet level air flow design for a rack of storage arrays.

Another type of storage rack level cooling architecture uses top of rack exhaust as shown below.

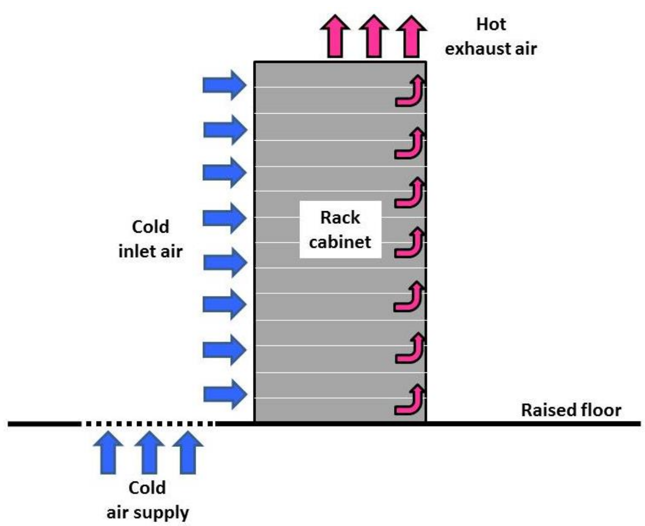

Figure 15 Cabinet level air flow design for storage arrays with top of rack exhaust.

A key point for the air flow design of any rack or cabinet is to avoid creating back pressure on the exhaust of the IT equipment. One example where this can occur is in a "chimney" design cabinet such as the one shown above in **Figure 15**. If the cabinet depth is too short, the air flow will have to make a very sharp right angle turn creating back pressure on the IT equipment. Back pressure is detrimental to equipment cooling and energy efficiency because the fans inside the IT equipment will have to spin faster and consume more energy to overcome the back pressure, for a given air inlet temperature.

ASHRAE recommends front to back cooling air flow architecture (shown in **Figure 12**,

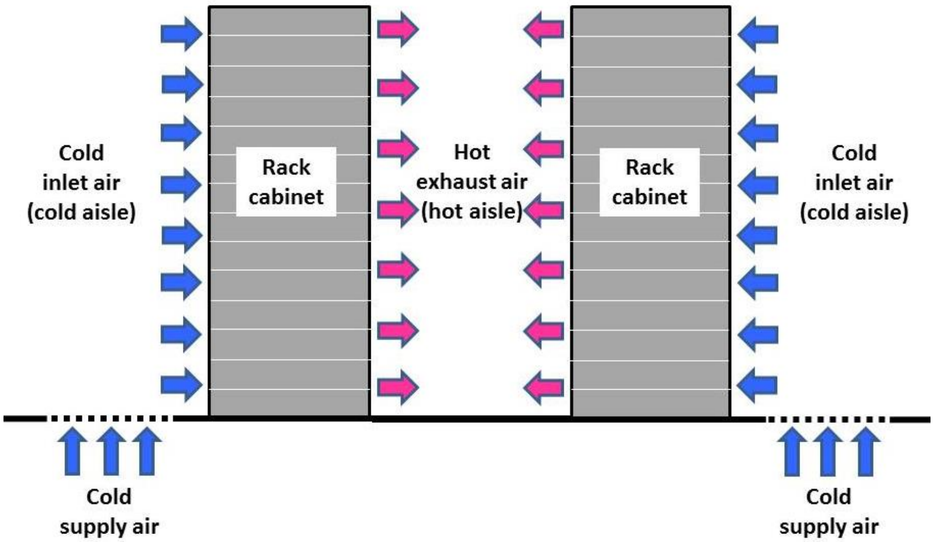

Figure 13, and **Figure 14**) for both individual storage array building blocks as well as for rack cabinet level storage arrays. Front to back cooling is recommended because it supports data center cooling best practices such as hot aisle/cold aisle layout and hot aisle/cold aisle containment.

Integrating storage arrays into a data center requires careful planning and design to meet the air flow requirements of the equipment. The amount of rack level air flow depends on a number of factors including: a) the power density of the equipment, and b) the temperature of the cooling air supply to the rack. Most storage equipment OEMs provide air flow requirements in the owner's manual or other customer facing documentation. Many of the OEM provided cooling air flow estimates are conservative and represent a worst case with the power consumption level the highest and the inlet air temperature the highest. The cooling air flow requirements for a piece of equipment are much more complex than just a single value. The rack level air flow needed to meet the cooling requirements will depend on the power dissipation of the equipment and the temperature of the inlet air. Most storage arrays employ a step function fan control algorithm that runs the fans at a very slow RPM to save energy during periods of normal operation when the IT equipment inlet cooling air stream is cool and then significantly increases the fan RPM in steps when more cooling is needed when internal power consumption levels are higher or when the temperature of the inlet air increases. An example of the rack level cooling air flow needed for a storage array is shown below.

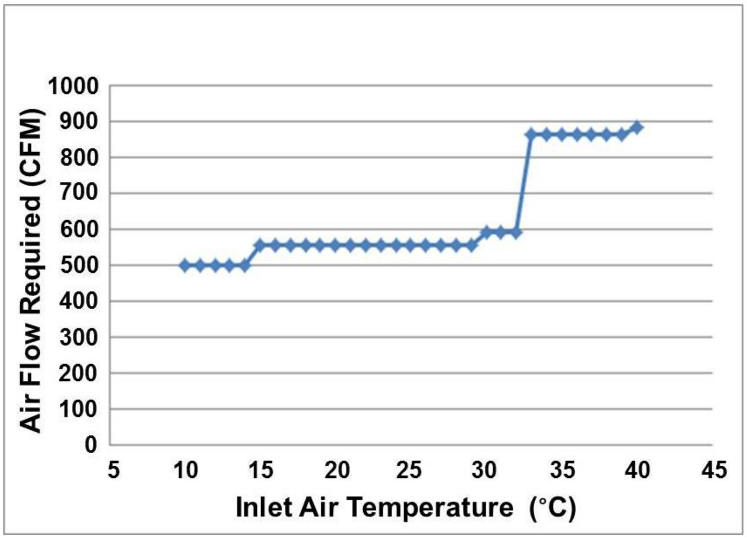

Figure 16 Example of rack level storage array air flow as a function of IT inlet air temperature.

In the figure above, the rack level air flow required shows only a gradual increase from 10 to 32°C and then a rapid increase at and above 33°C. **Figure 16** is just one example for one type of equipment. Adoption of more sophisticated fan control algorithms by OEMs may be an energy saving opportunity for storage arrays. For example, most servers are already using polynomial based fan algorithms to improve energy efficiency. Another reason storage array designers may want to consider a more sophisticated fan control is workloads can change the temperature of solid state drives much more rapidly than hard disk drives because solid state drives have a very small thermal mass compared to hard drives. This is especially true for PCIe SSDs which can have much higher power consumption than a typical hard drive. Temperature sensors should be positioned directly on or as close as possible to heat sources (vs. relying on air temperature measurements) so as to provide rapid and accurate feedback to the fan control system. More sophisticated fan control algorithms, such as polynomial, will generally deliver a faster and more accurate cooling response to work load driven temperature changes.

The fan curve behavior and rack level cooling air requirements may vary widely among different models and across suppliers. In the ASHRAE Thermal Guidelines book [13], ASHRAE recommends storage OEMs provide more granularity on the air flow requirements of their equipment in their customer facing documentation. Instead of providing a single worst case value for air flow, they should provide a matrix of air flow values as a function of equipment power consumption and inlet air temperature. Accurate air flow requirements are a very important input to data center design and operation and they are especially important for data centers that plan to use economization over a wide temperature range.

Most data centers contain dis-similar types of equipment with widely differing air flow requirements, power dissipation, and temperature control ranges. One approach is to cool the entire data center to within the limits of the equipment requiring the most restrictive environmental range, such as tape. However, this could result in over-cooling and over-controlling the entire data center to meet the needs of just one type of equipment. Most data centers are designed with multiple environmental control zones to help maintain a uniform data center temperature even when equipment heat loads and workloads are not necessarily uniform. Some data centers even have rooms that are completely isolated and separately controlled environmental zones. ASHRAE recommends the adoption of multiple data center environmental control zones as a best practice. Where possible, equipment with the same or similar environmental control requirements should be grouped together in the same zone or zones. For example, if all of the tape equipment was located in one zone it could be more closely controlled than the rest of the data center. A zone based approach to cooling and physically locating equipment will realize power and cooling cost savings for a majority of the data center while making sure critical equipment, such as tape or batteries, is controlled within narrower specified environmental limits. For more information on data center energy efficiency best practices, consult the ASHRAE Datacom Series books "Green Tips for Data Centers" and "Best Practices for Datacom Facility Energy Efficiency, Second Edition" [14], [15].

While the preceding discussion has been entirely about air cooling, it is worth noting there are a number of innovative cooling methods based on liquid. There are two broad types of liquid cooling: a) recirculating cooling systems with a pump and heat exchanger, and b) immersion cooling where the entire IT equipment is submerged in a bath of liquid such as mineral oil. A discussion of these cooling methods can be found in Appendix C.

2.3 Energy Saving Recommendations

Utilize Tiered Storage

Tiered storage involves mixing different environments within a single system. Various performance types (tiers) of hard drives coexist in a system and allow movement of data between tiers. Tiered storage enables consolidation of many arrays with different performance levels into one large array for more efficient use of resources and higher utilization. The hardware logic and power infrastructure required to run each of the arrays is more efficiently amortized over more drives. In general, there are 4 storage tiers each consisting of the following different drive types:

- Tier 0 – Solid state drives, SSD's,
- Tier 1 – 15k RPM Fibre Channel drives
- Tier 2 – 10k RPM Fibre Channel drives
- Tier 3 – 7.2k or 5k Fibre Channel or SATA Drives

Each drive type can be ideally selected and dedicated to run specific host applications and meet storage performance requirements. Tier 0 SSD's are an energy savings alternative for I/O intensive applications. One SSD can achieve the I/O rate of up to 30, 15k RPM drives and result in a 98% energy savings [14]. Older, unused data can be moved off the high performing drives to the slower, high capacity drives in Tier 3. This can save energy by reducing the number of high performance drives needed.

The different Tier levels can also be configured with different RAID levels. For instance, the high performance, critical applications can use a RAID 1 or mirrored configuration for higher availability. This requires twice as many drives for the same capacity, which is not efficient from an energy standpoint. Lower performing Tiers can use RAID 5 or 6, which require less drives and more useable drive capacity, thereby increasing energy efficiency.

Advantages of using tiered storage include:

1. Less number of physical drives and systems required
2. More efficient use of storage resources
3. Lower power and space consumption

Some of the disadvantages of tiered storage are:

1. Requires careful configuration design to optimize performance
2. Can reduce Bandwidth due to higher channel utilization
3. Could limit the number of host connections

Consolidate Platforms

Trading in older, less energy efficient storage arrays and consolidating into newer arrays can reduce power consumption and cooling requirements and also reduce footprint in the data center [14]. An enabler to consolidation is the ability to utilize larger capacity drives that have the same performance and power consumption, but can seamlessly replace many smaller capacity drives. Also, more efficient RAID levels that are available today, such as RAID 5 7+1, enable higher useable capacity and therefore less drives needed.

Advantages of platform consolidation are:

1. Upgrade and refresh technology
2. Fewer machines to power and maintain
3. Standardize on service and training requirements
4. Use replaced machines for backup, archiving and disaster recovery

Some disadvantages include:

1. Time consuming to migrate data off older array
2. Fewer larger capacity drives could reduce performance levels
3. Potential to run out of host connectivity.

"Pay As You Grow" Configuration Flexibility

It is more energy efficient to maintain a high utilization of storage which translates into less drives and hardware required [14]. Storage arrays can be purchased with minimum configurations to meet present needs and then scaled as storage and connectivity needs grow. When utilization is maximized and more storage is needed, individual drives or entire frames of drives can be purchased in efficient quantities to meet storage requirements. Virtual or thin provisioning can also help gain more useable capacity for the drives and thereby reduce the hardware requirements and help delay upgrades.

Advantages of "pay as you grow" configuration flexibility are:

1. Reduce initial Cap Ex
2. Reduce power consumption
3. Reduce foot print

Disadvantages of configuration flexibility include:

1. Requires advanced facilities planning for future upgrades

Performance Management

Utilize storage software to control and manage performance of storage arrays for energy efficiency [14]. One aspect that can be controlled is minimizing disc access and utilizing cache algorithms; this enables larger capacity discs to be used and yet still achieve high performance levels. Dynamic cache partitioning can provide priority to certain applications and throttle other applications based on policy. This can enable slow, high capacity drives to be used and still meet performance and service levels. Arrays can be further optimized for performance and efficiency with combination of slow, high capacity drives and Solid State drives, which can boost the I/O rate.

Advantages of performance management include:

1. Use slow, high capacity drives for energy efficiency
2. Reduce Cap Ex and footprint

Some disadvantages of performance management are:

1. SSD's can be expensive, but can be offset by replacing additional hardware.

Managed Capacity

This technique involves methods that reduce the amount of data that actually is stored to disc. One method is the use of de-duplication software that can prune and eliminate redundant copies of data and store incremental changes, thus reducing growth of storage, reducing hardware requirements and saving energy [14].

Another method is to employ an ILM (Information Lifecycle Management) strategy to segregate and disposition different types of data for improved energy efficiency. Examples of this include archiving older, un-accessed data and storing it on high capacity drives that use scheduled energy saving shutdown modes. For backup application, clones or snaps can be used to record incremental changes and not backup the entire database. Protocols can be established to sort older data and label it ready for delete. This can free up additional storage space and prevent having to add new storage.

Some advantages of managed capacity include:

1. Reduce the amount of data being stored to disc
2. Reducing data also reduces downstream storage such as backup and archiving
3. Reduce the number of systems powered up waiting for a fail over event

Disadvantages of managed capacity are:

1. De-duplication software can slow system down if not run in the background
2. Data that is refreshed frequently will not benefit much from de-duplication

3 Environmental Specifications

3.1 Review of Industry Environmental Specifications

The most commonly used environmental thermal guidelines and specifications for storage equipment are the ASHRAE thermal guidelines [16]. Network Equipment Building Standard (NEBS) [17] and European Telecommunications Standards Institute (ETSI) [18] are telecom specifications and are included for comparison. It should be noted the telecommunications industry does use a small amount of telecom rated storage equipment. A summary of the ASHRAE and telecom environmental classes is given below in **Table 3**, **Table 4**,

Table 5, and **Table 6**.

Table 3 Summary of ASHRAE 2011 Thermal Guideline Classes [16].

Classes (a)	Equipment Environmental Specifications for Air Cooling							
	Product Operations[b,c]					Product Power Off[c,d]		
	Dry-Bulb Temperature (˚C)[e,g]	Humidity Range, non-Condensing[h,i]	Maximum Dew Point (˚C)	Maximum Elevation[e,j] (m)	Maximum Temperature Change in an Hour (˚C)[f]	Dry-Bulb Temperature (˚C)	Relative Humidity (%)	Maximum Dew Point (˚C)
Recommended* (Suitable for all 4 classes; explore data center metrics in this paper for conditions outside this range)								
A1 to A4	18 to 27	5.5ºC DP to 60% RH and 15ºC DP						
Allowable								
A1	15 to 32	20% to 80% RH	17	3050	5/20	5 to 45	8 to 80	27
A2	10 to 35	20% to 80% RH	21	3050	5/20	5 to 45	8 to 80	27
A3	5 to 40	-12˚C DP & 8% RH to 85% RH	24	3050	5/20	5 to 45	8 to 85	27
A4	5 to 45	-12˚C DP & 8% RH to 90% RH	24	3050	5/20	5 to 45	8 to 90	27
B	5 to 35	8% RH to 80% RH	28	3050	NA	5 to 45	8 to 80	29
C	5 to 40	8% RH to 80% RH	28	3050	NA	5 to 45	8 to 80	29

*The 2008 recommended ranges as shown here can still be used for data centers.

Notes for Table 2.3, 2011 Thermal Guidelines – SI Version

a. Classes A1, A2, B, and C are identical to 2008 classes 1, 2, 3 and 4. These classes have simply been renamed to avoid confusion with classes A1 thru A4. The recommended envelope is identical to that published in the 2008 version of Thermal Guidelines.

b. Product equipment is powered ON.

c. Tape products require a stable and more restrictive environment (similar to Class A1). Typical requirements: minimum temperature is 15°C, maximum temperature is 32°C, minimum relative humidity is 20%, maximum relative humidity is 80%, maximum dew point is 22°C, temperature change requirement is less than 5°C in an hour, the change in humidity is less than 5% RH in an hour with no condensation.

d. Product equipment is removed from original shipping container and installed but not in use, e.g., during repair maintenance, or upgrade.

e. Classes A1, A2, B and C - Derate maximum allowable dry-bulb temperature 1°C/300 m above 900 m. Above 2400 m altitude, the derated dry-bulb temperature takes precedence over the recommended temperature. A3 - Derate maximum allowable dry-bulb temperature 1°C/175 m above 900 m. A4 - Derate maximum allowable dry-bulb temperature 1°C/125 m above 900 m.

f. For tape storage equipment: 5°C in an hour. For all other IT equipment: 20°C in an hour and no more than 5°C in any 15 minute period of time. The temperature change of the IT Equipment must meet the limits shown in the table above, and is calculated to be the maximum air inlet temperature minus the minimum air inlet temperature within any time window as specified. The 5°C and 20°C temperature change is considered to be a temperature change within a specified period of time and not a rate of change.

g. With diskette in the drive, the minimum temperature is 10°C
h. The minimum humidity level for class A3 and A4 is the higher (more moisture) of the -12°C dew point and the 8% RH. These intersect at approximately 25°C. Below this intersection (~25°C) the dew point (-12°C) represents the minimum moisture level, while above it, RH (8%) is the minimum.
i. Moisture levels lower than 0.5˚C DP, but not lower -10˚C DP or 8% RH, can be accepted if appropriate control measures are implemented to limit the generation of static electricity on personnel and equipment in the data center. All personnel and mobile furnishings/equipment must be connected to ground via an appropriate static control system. The following items are considered the minimum requirements (see Appendix A for additional details):
   1) Conductive Materials
      a) Conductive flooring
      b) Conductive footwear on all personnel that go into the datacenter, including visitors just passing through;
      c) All mobile furnishing/equipment will be made of conductive or static dissipative materials.
   2) During maintenance on any hardware, a properly functioning wrist strap must be used by any personnel who contacts ITE.
j. To accommodate rounding when converting between SI and I-P units, the maximum elevation is considered to have a variation of +/- 0.1%. The impact on ITE thermal performance within this variation range is negligible and enables the use of rounded values of 3050 m (10,000 ft).

New note: operation above 3050 m requires consultation with IT supplier for each specific piece of equipment.

Table 4 Summary of NEBS equipment aisle[1] air temperature and humidity limits [17].

Conditions	Limits
Temperature	
• Operating (up to 1829m [6000ft])	5 to 40°C
• Short-term[2]	-5 to 50°C
• Short-term with fan failure	-5 to 40°C
Rates of Temperature Change	
• Operating	30°C/hour
Relative Humidity	
• Operating	5 to 85%
• Short-term[2]	5 to 93% but not to exceed 0.026 kg water/kg of dry air

Notes:
1. Equipment aisle refers to conditions at a location 1524 mm (60 in) above the floor and 381 mm (15.0 in) in front of the equipment. Equipment test temperatures are defined in Section 5.1, "Temperature, Humidity, and Altitude Test Methods," based on equipment configuration (frame-level or shelf-level) and air-inlet location.
2. Short-term refers to a period of not greater than 96 consecutive hours, and a total of greater than 15 days in 1 year. (This refers to a total of 360 hours in any given year, but not greater than 15 occurrences during that 1-year period.)

Table 5 Summary of ETSI Class 3.1 and 3.1e environmental requirements [18]**.**

	Continuous Operation	Class 3.1: ≤10% of Operational Hours	Class 3.1e: ≤1% of Operational Hours
Temperature Ranges	10 to 35°C	5 to 10°C, and 35 to 40°C	-5 to 5°C, and 40 to 45°C
Humidity Ranges	10% to 80%RH[1]	5 to 10%RH[2], and 80 to 85%RH[3]	5 to 10%RH[2], and 85 to 90%RH[3]

[1]With minimum absolute humidity of no less than 1.5 g/m^3 and a maximum absolute humidity of no more than 20g/m^3

[2]With minimum absolute humidity of no less than 1 g/m^3.

[3]With maximum absolute humidity of no more than 25 g/m^3.

Note: maximum rate of temperature change for continuous operation, Class 3.1 and Class 3.1e is 0.5°C/minute averaged over 5 minutes.

Table 6 Summary of maximum altitude ratings and de-rating values [16]**.**

Specification & Class	Minimum Altitude	Maximum Altitude	De-Rating (Metric Units)	De-Rating (I-P Units)
ASHRAE - Class A1	NA	3050m (10000ft)	1°C/300m above 950m	1.8°F/984ft above 3117ft
ASHRAE – Class A2	NA	3050m (10000ft)	1°C/300m above 950m	1.8°F/984ft above 3117ft
ASHRAE – Class A3	NA	3050m (10000ft)	1°C/175m above 950m	1.8°F/574ft above 3117ft
ASHRAE – Class A4	NA	3050m (10000ft)	1°C/125m above 950m	1.8°F/410ft above 3117ft
NEBS GR-63-CORE	NA	3960m (13000ft)	1°C/305m above 1829m	2.2°F/1000ft above 6000ft
ETSI – Classes 3.1, 3.1e	70kPa (approx. 381m (1250ft) below sea level	106kPa (approx. 3048m (10000ft) above sea level	NA	NA

If one compares and contrasts ASHRAE, NEBS and ETSI, the ASHRAE specifications are written only for continuous 7 days x 24 hours operation whereas the NEBS and ETSI specifications allow for limited time temperature excursions outside of the continuous operation range.

3.2 Storage Equipment Environmental Ratings

A benchmarking survey of 8 of the major storage array suppliers and more than 200 storage array models was carried out to understand the maximum operational temperature ratings of their offerings as shown below in **Figure 17**.

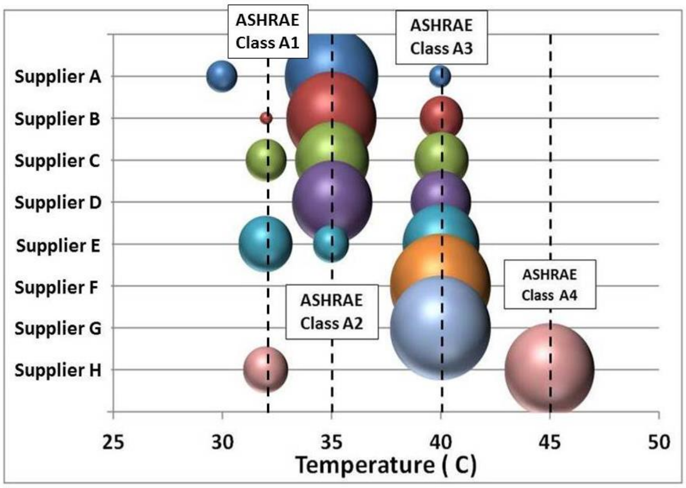

Figure 17 OEM benchmarking survey of maximum operational rated air inlet temperatures.

Each row along the vertical axis corresponds to a different supplier. For each supplier the relative area of the bubbles represents the percentage of their portfolio with that maximum rated temperature. The sum of the area of all of the bubbles along a row is 100%.

The ASHRAE Classes in **Table 3** are structured such that Class A1 represents typical older vintage IT equipment. The conditions of class A2 meet or exceed those of Class A1. Class A3 meets or exceeds the conditions of Classes A2 and A3. Therefore storage array products are, in general, able to handle temperatures higher than typical Class A1 equipment. The results in **Figure 17** show most commercially available storage arrays either meet an ASHRAE Class A2 (35°C) rating or an ASHRAE Class A3 (40°C) rating. This is fairly well aligned with the maximum temperature ratings of recent vintage servers.

Non-operational specifications are also an important part of the environmental specifications for IT equipment including storage arrays. Non-operational conditions include shipping, long term storage, and any time the system may spend deployed in a data center but powered down. Benchmarking of non-operational specifications was conducted for the same suppliers and products used in **Figure 17** as shown below in **Figure 18** and **Figure 19**.

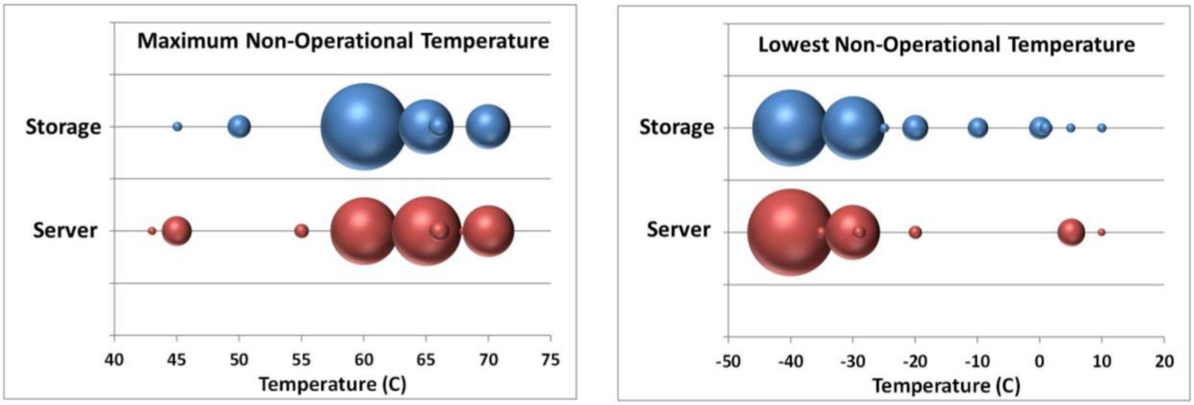

Figure 18 Maximum and minimum non-operational temperature ratings for disk based storage arrays and servers.

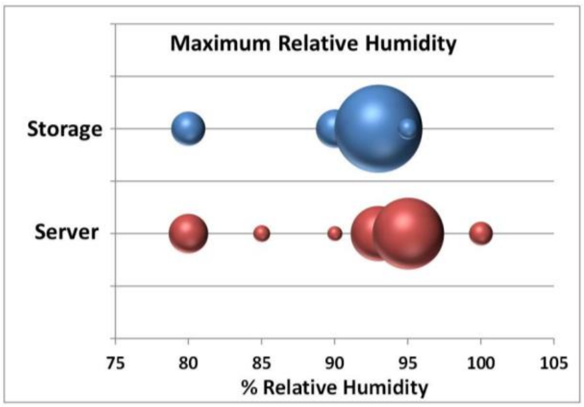

Figure 19 Maximum non-operational relative humidity for storage arrays and servers.

For a storage array, typical non-operational conditions are something like -40 to 60°C with a maximum relative humidity of 93%. However, the verbiage that accompanies most storage array environmental specifications doesn't differentiate between shipping, long term storage or powered down deployment. In fact, most specifications have a "one size fits all" set of non-operational conditions and some don't even include a maximum dew point limit. Another interesting point to note is most of these specifications place no limit on the amount of time that can be spent in non-operational condition.

Typical non-operational storage array specs are -40 to 60°C and 5 to 93% relative humidity with a 33°C maximum dew point. These conditions are much wider than Class A2 operational conditions which are 10 to 35°C 20 to 80%RH with a 21°C maximum dew point. Typical hard disk drive operational and non-operational specifications are very similar to the specifications for storage arrays and, at first glance, the two sets of specifications appear to be aligned. However, most hard disk drives are not hermetically sealed and the environment inside the drive is affected by conditions outside the drive. Hard drives have a small vent hole that allows the environment inside and outside the hard drive to slowly equilibrate. Hard drives can tolerate exposure to upper limit non-operational temperature and humidity conditions but only for a limited period of time before the drive suffers permanent degradation that could lead to failure. Current hard disk drive specifications don't capture these important time limitations. Thus, it may be possible for an end user to ship and store their storage arrays within the manufacturer's specification and have significant hard drive damage or even hard drive failures. At present the environmental capability of the hard disk drives and the storage array non-operational specifications are not in alignment because time limits on exposure to the maximum non-operational conditions, especially high humidity, are not defined. It should be noted that the effect of low humidity on hard disk drive reliability is not well understood and long term exposure to lower limit humidity values should also be avoided.

Hard drives in their original factory packaging, consisting of a sealed bag with desiccant, can withstand high humidity levels for prolonged periods of time because the drive is insulated from the external environment. Most storage arrays are shipped with the drives removed from their factory packaging and already installed in the drive bays of the storage array chassis where the drives have no protection from extreme environmental conditions.

To address the need for time limits on non-operational specifications, ASHRAE recommends dividing the non-operational state in two: shipping and long term storage. Shipping is a wider set of excursion based environmental conditions that are allowed for only a limited period of time. Shipping is typically an uncontrolled environment such as sea shipment inside a steel shipping container, shipment in the back of a truck, or shipment in the cargo hold of an airplane. Long term storage is a narrower environmental window that is allowed for an extended period of time and is typically a controlled environment. The complete ASHRAE recommendation is given below.

Table 7 ASHRAE recommended HDD and HDD based IT equipment non-operational and operational conditions as well as operational conditions for HDDs alone.

		Temperature (°C)		Humidity (%RH)[3]		Dew Point (°C)	
	Time Limit	Lower Limit	Upper Limit	Lower Limit	Upper Limit	Maximum	Comments
Non-Operational - System							
Shipping[1]							
Baseline	No Limit	5°C	45°C	8%	90%	24°C	ASHRAE Class A4
Excursion	14 days	-40°C	65°C	5%	95%	33°C	
Long Term Storage[2]							
Allowable	Limited[4]	5°C	45°C	8%	90%	24°C	ASHRAE Class A4
Recommended	Limited[4]	18°C	27°C	5.5°C DP	60%	15°C	ASHRAE recommended range
Operational - System							
ASHRAE Class A1	No Limit	15°C	32°C	20%	80%	17°C	
ASHRAE Class A2	No Limit	10°C	35°C	20%	80%	21°C	
ASHRAE Class A3	No Limit	5°C	40°C	8%[5]	85%	24°C	
ASHRAE Class A4	No Limit	5°C	45°C	8%[5]	90%	24°C	
Operational – Hard Disk Drive Only							
Allowable	No Limit	5°C[6]	60°C[6]	5%	90%	24°C	

[1] Shipping assumes either: a) the system being shipped is in factory sealed shipping container and the drives have no end user data on them, or b) the hard drive is in a factory sealed plastic bag with no customer or end user data stored on it.
[2] Long term storage applies to both storage arrays as well as individual hard drives. No special packaging of the system or the drives is assumed. The drives may have end user data stored on them.
[3] Ambient air conditions must be non-condensing at all times.
[4] Hard disk drives are not intended long term or archival storage of data in a non-operational condition. For non-operational storage longer than approximately 2 years, the data should be migrated to an archival storage media such as tape.
[5] Along with a -12°C minimum dew point
[6] Temperature is that of the air inside the hard disk drive chassis as measured by the temperature sensor internal to the drive.

Table 8 ASHRAE recommended SSD and SSD based IT equipment non-operational and non-operational conditions as well as operational conditions for the SSD alone.

		Temperature (°C)		Humidity (%RH)[4]		Dew Point (°C)	
	Time Limit	Lower Limit	Upper Limit	Lower Limit	Upper Limit	Maximum	Comments
Non-Operational - System							
Shipping[1]							
Baseline	3 months	5°C	40°C	8%	90%	24°C	
Excursion	14 days	-40°C	65°C	5%	95%	33°C	
Long Term Storage[2]							
Allowable[3]	3 months	5°C	40°C	8%	90%	24°C	
Recommended	1 year	18°C	27°C	5.5°C DP	60%	15°C	ASHRAE recommended range
Operational - System							
ASHRAE Class A1	No Limit	15°C	32°C	20%	80%	17°C	
ASHRAE Class A2	No Limit	10°C	35°C	20%	80%	21°C	
ASHRAE Class A3	No Limit	5°C	40°C	8%[5]	85%	24°C	
ASHRAE Class A4	No Limit	5°C	45°C	8%[5]	90%	24°C	
Operational – Solid State Drive Only							
Allowable	No Limit	5°C[6]	60°C[6]	5%	90%	24°C	

[1]The SSD can tolerate higher temperatures but if the drive is shipped or stored at temperatures above 40°C then the data retention is reduced to something less than 3 months.
[2]Drive back-up should always be carried out to guard against data loss. Use of SSD media for archiving of data is not recommended.
[3]3 month data retention at 40°C per JEDEC enterprise class SSD retention use requirements [19]
[4]Ambient air conditions must be non-condensing at all times.
[5]Along with a -12°C minimum dew point
[6]Temperature is that of the SSD chassis.

3.3 Environmental Conditions for Tape Media

Tape is a very unique type of storage in that it has one set of environmental conditions for the hardware and another for the media. Unlike other forms of storage, tape media cartridges are easily removed and are sometimes archived at a different physical location. The tape drives themselves do not contain data. Only the tape cartridges store data which explains why their storage conditions are more critical. Each drive and tape manufacturer has slightly different environmental operating and storage recommendations. A standardized recommendation of environmental conditions for tape hardware and tape media is given below in

Table 9.

Table 9 Tape Media, Tape Drive, and Tape Library Hardware Recommendations.

		Temperature (°C)		Humidity (%RH)*		Wet Bulb (°C)	
	Time Limit	Lower Limit	Upper Limit	Lower Limit	Upper Limit	Max.	Comments
Tape Drives & Library Hardware: Non-Operational							
Shipping Excursion	10 days	-40°C	60°C	10%	95%	29°C	Does not include tape cartridges with data
Shipping and Long Term Storage	No Limit	10°C	40°C	10%	95%	35°C	Check manufacturer spec
Tape Drives & Library Hardware: Operational							
Allowable	No Limit	10°C	40°C	10%	80%	26°C	Check manufacturer spec
Recommended	No Limit	15°C	32°C	20%	80%	26°C	
Tape Media Cartridges							
Shipping Excursion	10 days	-23°C	49°C	5%	80%	26°C	Cartridges come from the manufacturer pre-written with servo patterns
Allowable Shipping and Long Term Storage	No Limit	10°C	32°C	20%	80%	26°C	Check manufacturer specs
Recommended Long Term Storage for Archive	No Limit (up to 30 years)	15°C	25°C	15%	50%	26°C	

*Ambient air conditions must be non-condensing at all times. Per the 2011 ASHRAE Thermal Guidelines, the humidity rate of change must be <5%RH/hour for tape media, tape drives, and tape libraries.

Per the recommendation in

Table 9 above, tape drives and libraries with tape media are not suitable for ASHRAE Class A3 and A4 environments. The humidity ranges of Classes A3 and A4 exceed the allowable ranges for tape. Use of tape beyond the ASHRAE recommended ranges above may result in a higher failure rate and a higher data error rate.

3.4 Gaseous Pollution and Dust

Management of gaseous pollution and dust are important to the long reliability of storage IT equipment. The ASHRAE recommendation [20] is to maintain an ISA-71 G1 level of corrosion [21] which corresponds to <200A/month silver corrosion rate and <300A/month copper corrosion rate. A detailed discussion of gaseous pollution, dust filtration requirements, and corrosion is beyond the scope of this document but can be found in the ASHRAE book "Particulate and Gaseous Contamination in Datacom Environments" [20].

Since the conditions that can impact corrosion rates are dynamic, the corrosion level in any data center environment should be measured using corrosion coupons on a periodic basis or with a real-time corrosion rate monitoring system for compliance to the ASHRAE guideline of ISA-71 G1. Corrosion rate is a complex and synergistic combination of gaseous pollution levels and relative humidity. Pollution and humidity levels can vary significantly over the seasons of a year. Gaseous contaminants are generally invisible; therefore data center environments that appear to be "clean" can still have corrosion rates well above G1. Data centers that use a large fraction of make-up air from the outside, such as with air-side economization, should monitor their corrosion rates closely especially if the data centers are in close proximity to pollutant sources such as a major highway, agricultural crop spraying, or being downwind of a polluter such as a coal fired power plant or a tire and rubber factory.

Provided ASHRAE air filtration requirements for incoming and recirculated air are maintained (ISO Class 8 per ISO 14644-1 with a 95% upper confidence limit), dust is not generally a problem in most data centers. However, there are some important exceptions. Activities that take place inside the data center, such as drilling and drywall construction, can generate large amounts of particles in a localized area. The environment of IT equipment near these activities can easily exceed ASHRAE air quality standards for particles. In the case of drywall construction, gypsum dust is especially damaging to tape and tape heads. IT equipment should be protected by installing temporary partitions, such as plastic sheeting, around construction areas to make sure none of the particles being generated reach any of the IT equipment. After any drilling or construction is finished, residual particles should be carefully wiped or vacuumed away to prevent them from being stirred up into the IT equipment cooling air stream. Particle based fire suppression systems are not recommended for data centers containing tape storage equipment.

4 Reliability

4.1 Hard Disk Drive Reliability

HDD reliability is the most critical factor to data integrity and data availability. Among all the components and failure sources, the head-disk interface (HDI) is a critical area which is responsible for the most common HDD failure mechanisms. The HDI is constructed in nm (nanometer) order, and the HDD has been said to be one of the very first mechatronics that realized the use of nano-technology in high volume production. In these fields, the nano-mechatronics plays an important role; processing in nanometer size, making a move in nanometer length, and controlling in nanometer order [22]. A cut away photo of the internal parts of a hard disk drive is shown below in Figure 20. The HDI is the nm sized gap between the head at the end of the slider and the magnetic media on the platters.

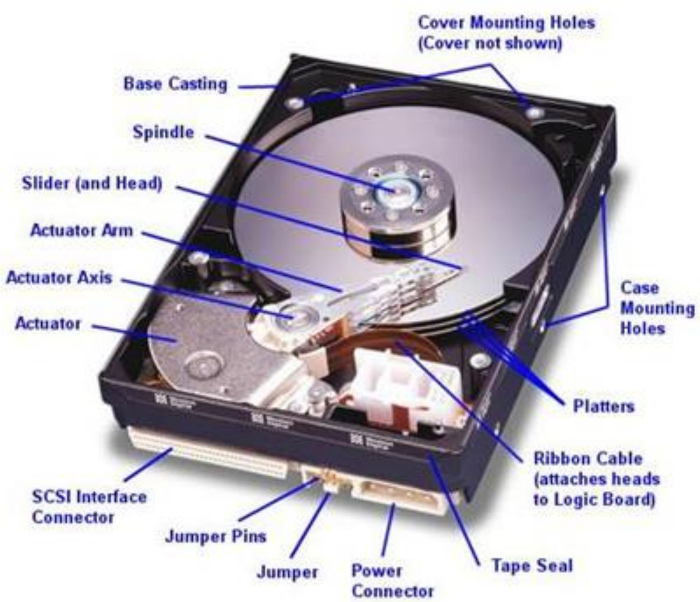

Figure 20 Cut away view of the parts inside a hard disk drive.

HDD reliability has been improved substantially over the last two decades. In early 1990's, most manufacturers specified MTTF between 100,000 and 250,000 hours. However, by early 2000's, many HDD suppliers indicate MTTF in the neighborhood of 400,000 to 600,000 hours for Desktop and Mobile product and 1 - 1.2 million hours for Enterprise product. Now Enterprise product MTTF is as high as 2 million hours. Yet the HDD is still a significant weak link in the reliability of data storage because storage arrays use so many of them. To put these MTBF numbers in perspective, 100,000 hours is 11.4 years. MTBF is not a "lifetime" at which failures start to occur. Rather it is the average time the failures occur. Because it is an average, some failures occur much earlier than the MTBF value and some occur much later. The early failures are the ones that have the most detrimental impact on the reliability of storage arrays.

According to one source [23] disk drive failures are responsible for 20-55% of storage subsystem failures. Also, these storage sub-system failures have a tendency to be self-correlated and some failures occur in bursts rather than being spread out randomly over time [23]. Because storage arrays are comprised of a large number of disks, the impact of a single disk drive failure needs to be managed and mitigated by the implementation of RAID. There are two primary categories of disk drive failure: a) those that leave the HDD working but corrupt the data, and b) failure of the entire HDD.

Data corruption

Bit error rate (BER) is a statistical measure of the effectiveness of the drive systems (electrical, mechanical, firmware) to work together to write or read data. The most problematic bit errors are the ones where the data is missing or corrupted without the knowledge of the user. These are sometimes called latent defects because they are present but not yet detected. Latent defects can be caused by data that wasn't written well or was erased or corrupted after being written correctly. Some causes of latent defects include excessive vibration (e.g. bumping or banging an HDD during a write or heavy walking across a poorly supported raised floor), a persistent increase in fly height caused by buildup of lubrication on the surface of the slider as well as thermal asperities, corrosion, scratches and smears on the drive media [24].

Potential Impact of Noise and Vibration on Disk Performance

There have been several industry reports [25], [26], and [27] of high noise levels, such as fire alarm horns and inert gas release from fire suppression systems, reducing the performance of hard disk drives and possibly causing data corruption. Fire alarm horns and suppression systems can generate noise levels in the range of 90 – 120db. Noise levels in and above this range can be detrimental to IT equipment because they generate a pressure wave which causes the HDD to fail. Vibration can, in some cases, compromise performance in spinning media such as hard disk drives. Some of the steps a data center can take to minimize the possibility of hard drive performance impact are:

- Avoid exposing sensitive IT equipment to noise levels above 110db.
- Locate fire alarm sirens and fire suppression system nozzles so they don't radiate noise directly at sensitive IT equipment containing hard drives, such as storage arrays.
- For inert gas fire suppression systems, install noise reduction gas release nozzles.
- Use enterprise quality hard disk drives that may be less sensitive to noise and vibration or use solid state drives.
- Use redundant data storage architectures such as RAID.
- Perform routine inspections and maintenance of all raised floor systems.

- Where possible, minimize external sources of vibration include activities such as railroad trains, construction and heavy equipment in close proximity to the data center.
- Minimize internal sources of vibration such as air handlers, generators, construction, and movement of heavy equipment across the floor of the data center.

It should be noted that IT equipment racks and cabinets usually absorb some vibration so the level of vibration transmitted to the IT equipment may be less than the vibration levels transmitted to the floor of the data center.

The threshold at which vibration is detrimental to hard drives and storage arrays is highly dependent many factors some of which are the type of rack, the design of the enclosure of the storage array, the type of hard disk drive, and the technology being used in the disk drive. Because there are so many factors involved, ASHRAE does not have a recommendation on specific vibration limits. If vibration is suspected as a possible cause of hard disk performance degradation or failures, accelerometers should be installed so performance degradation and/or failures can be correlated to vibration levels. The development of an industry standard to define acceptable data center vibration levels and monitoring of those levels is an opportunity for future work.

Disk Failure

There are a number of types of operational failures that could cause an entire hard disk drive to fail. Some of these include: a) bad servo track, b) can't stay on track, c) SMART limits exceeded, and d) changes in the magnetic properties of the head. Hard disk drive operational failure rates can vary widely between manufacturers, models and manufacturing vintage. Factors responsible for this variation include: a) airborne contamination (particles) inside the drive, b) manufacturing defects and variability, c) design changes – while most of these are beneficial to performance, some inadvertently degrade reliability, d) changes to the drive manufacturing process to improve yield may degrade reliability, and e) variability in end user usage patterns. In general, most disk failures are caused by disk drive manufacturing defects and design issues. These types of failures are generally not preventable by the user. A detailed explanation of disk failure mechanisms can be found in Appendix D.

Potential Impact of Sustained High Humidity

Most hard disk drives are not hermetically sealed – they have a small vent hole in the case of the drive to equilibrate atmospheric pressure between the inside and outside of the drive. The transfer of humidity through the vent hole is very slow, i.e. on the order of days or even weeks. Over the course of many days it is possible for enough moisture to seep through the vent hole to significantly change the humidity internal to the drive. Increased humidity can have the effect of decreasing the fly height of the head [28]. This occurs because compression under the head can result in localized super-saturation of water vapor. The drive depends on an air bearing surface to maintain the correct head fly height above the media. The air bearing surface doesn't provide as much lift when the air is super-saturated with water vapor thus decreasing the fly height of the head and increasing the probability of a performance impact or a failure due to a collision between the head and the media.

In general, high humidity excursions of no more than a few days don't last long enough to significantly change the humidity level inside the hard drive. However, sustained exposure to high humidity levels longer than the non-operational values in

Table 7 should be avoided. For environments where humidity may be a problem, some hard disk drive manufacturers now offer helium filled hermetically sealed drives. These drives may also be useful for applications where a vent hole would not be desirable such as oil immersion cooling.

4.2 Solid State Drive Reliability

Solid state drives (SSDs) based on Negated AND (NAND) flash memory are a relatively new type of storage media. The amount of reliability data from large deployments of SSDs over an extended period of time (years) is still somewhat limited. The reliability of SSDs is largely determined by the reliability of the flash memory and the controller circuitry that supports it.

A detailed discussion of flash memory cells, cell degradation mechanisms, and management of flash media degradation at the drive level can be found in Appendix E.

One feature of solid state drives is they tend to undergo a gradual failure of the media, giving the user ample warning a drive replacement is needed before there is any risk of data loss. In practice, SSDs aren't yet reaching the wear out limit of the media. However, there are still scenarios where a failure of the circuitry that controls the flash memory array can cause a catastrophic drive failure resulting in data loss.

The preliminary field reliability data for SSDs is encouraging – they appear to have at least comparable reliability to hard disk drives. Also, there is no evidence that the SSDs that have been deployed have reached the program-erase cycling limit of the flash media. However, we still don't have a long history on large populations of SSDs to confirm this.

Solid state drives are currently selling for a premium to hard disk drives, for the same amount of storage. However, the advantage of a solid state drive is a higher I/O at lower than one could achieve with hard drives alone. Another advantage of SSDs is, for the same I/O, if a single SSD could replace 4 HDDs there would be a power and cooling cost savings as well as a net reliability improvement because there would be only one drive instead of four.

A key difference between SSD and HDDs is a temperature dependence of the degradation of flash media used in SSDs. Many SSDs have similar environmental specs to HDDs mainly because they are going into the same drive bays and many SSD suppliers have leveraged existing HDD specifications. Typical HDD and SSD operational temperature specifications are 5 to 60^{o}C. The temperature of the recording media typically scales with the inlet air temperature of the storage array. However, flash actually degrades more slowly as the temperature of the media increases. Flash media has a self-healing mechanism that is more efficient with increasing media temperature. The net effect of using an SSD at higher air inlet and media temperatures is still a gradual degradation of the media – the degradation occurs more slowly with increasing temperature. However, the higher media temperatures that aide in reducing write-erase cycling damage can negatively impact other areas including:

- Increase the failure rate of on-flash circuitry - the higher media temperatures that help the flash array could be slightly offset by a higher failure rate of the non-flash controller circuitry
- Shorten the intrinsic reliability lifetime of the flash array – higher media temperatures will also adversely impact the non-defect driven wear out (intrinsic reliability) of the flash media and other components
- Compromise charge storage devices such as super caps – super caps are a unique and highly specialized type of capacitor. They are typically rated to a maximum of 60-70°C for a limited period of time. These components are often used to assure data backup in the event of a power loss.

Application of SSDs at higher operating temperatures than HDDs may be possible but should be undertaken only after careful analysis of all of the factors listed above. If SSDs can be engineered to operate at higher temperatures than conventional HDDs but with comparable or better reliability, this would give storage array designers an extra degree of freedom and could enable more three dimensional stacking of drives such as drive behind a drive and embedding drives within a chassis.

Thermal throttling is a relatively new feature offered with some SSD devices. It provides another way to manage large temperature excursions that may occur within the SSD subsystem environment. The most common source of increased temperature within an SSD is controller-induced heating, often a result of a high concentration of write operations. Thermal throttling schemes can scale back available write performance slightly when a target temperature is reached at the SSD thermal sensor. A slightly higher target temperature can introduce a deeper throttling level in order to avoid a warranty temperature limit for example.

Another important consideration in the application of solid state drives is power-off data retention. The current JEDEC Solid State Technology Association (formerly known as the Joint Electron Device Engineering Council) enterprise SSD power-off data retention specification is 3 months [29] duration at 40°C at the maximum program-erase cycle count. If the drive was exposed to a temperature above 40°C during power-off, the data retention would be reduced to something less than 3 months. Conversely, if the drive had less than its maximum program-erase cycle count of wear it may be able to withstand a higher temperature than 40°C without data loss. Consult your SSD supplier for power off data retention specifications and advice on de-rating or up-rating these specifications based on your application. The JEDEC time and temperature limit should also be observed when the hardware is powered down such as when it is being moved from one physical location to another. Data back-up should always be carried out as a precaution against data corruption or loss.

Flash data retention is also an important factor that determines the shelf life of a new SSD prior to actual usage (i.e. shelf life). SSDs contain metadata – metadata is the information stored in flash to identify memory blocks that have been designated as bad from the supplier manufacturing test process. If the metadata is corrupted or compromised, the SSD may or may not function correctly. The storage temperature and duration of SSD hardware prior to deployment should be managed carefully to avoid the possibility of affecting the integrity of the metadata.

4.3 Battery Reliability

Batteries are a critical component of storage equipment. Batteries are used to keep system time of day and in applications such as memory cache backup used within Redundant Array of Independent Disks (RAID) adaptor card assemblies. In data centers large arrays of batteries are used as part of the UPS system to ensure a temporary alternate power supply to the IT equipment in case of power outage to allow systems to safely shut down and to provide time for backup power generators to come online. The discussion in this section will be confined to batteries installed within storage arrays or on storage related PCBAs such as RAID cards.

Loss of power at the system or data center level leading to abrupt system shut down and potential loss of data is highly undesirable. The use of batteries at the system level helps safeguard against this by powering the necessary hardware to either allow data to be transferred from volatile memory to non-volatile storage or to maintain the data in the volatile memory until power is restored.

Many design requirements play into the type of battery technology that is chosen to reside in the IT equipment, rack or data center. Some key requirements include energy density, charge and discharge rates, thermal environment, implementation of fuel gauges and microcode, and safety/environmental aspects. Summary tables of battery technologies along with their respective advantages and disadvantages can be found in Appendix F.

The environment in which the battery resides must be carefully managed to ensure that the battery life and behavior is as expected. In general, hot environments, batteries suffer from reduced life which can result in a battery fail and loss of data in case of an outage. At cold temperatures batteries can suffer from reduced capacity which can manifest as reduced performance as firmware limits the amount written to cache. Thus both temperature extremes should be avoided where possible. In general, for batteries in storage equipment, air ambient operating temperatures between 0ºC and 45ºC will not create a significant difference in battery performance. Storage equipment should be designed to take into account reduced battery capacity and life at temperature extremes where the battery temperature goes beyond this range.

As a rule-of-thumb the battery life of sealed lead acid batteries degrades by about 50% for every 7-10ºC increase in temperature [30]. Similarly, **Table 10** shows the impact of temperature on the percentage capacity degradation of a primary Li cell used in a server for the case where the battery degradation is only due to self-discharge (i.e., no other battery load current to the application). Assuming constant 10ºC uplift from the server inlet to the battery within the system, the table predicts that the expected battery life, where end of life is defined as 90% capacity remaining, will reduce from 3 years to 1.25 years when increasing from a steady 27ºC server inlet (maximum on the ASHRAE recommended envelope) to 40ºC server inlet (maximum on the ASHRAE Class A3 allowable envelope). The expected battery life increases if additional battery capacity degradation is allowed. For example, if end of life is defined as 80% capacity remaining, the expected life under the same conditions as defined above will become 6 years at 27ºC to 2.5 years at 40ºC server inlet. Of course, if the battery is providing load current in addition to the self-discharge it will further reduce the expected life. The additional application load current must be accounted for in the expected life calculation for those cases. In some cases, the equipment firmware may allow the user to monitor battery capacity and life. If the battery health cannot be monitored then the battery should be replaced periodically on a schedule based on its operating temperature.

Table 10 Expected capacity degradation from self-discharge as a function of temperature for a CR type Lithium Primary (Lithium Manganese Dioxide) coin cell battery commonly used in IT equipment [30].

Temperature (°C)	Capacity Degradation (% per year)
20	1
30	2
40	4
50	8
60	16

Humidity must also be carefully managed to ensure that corrosion effects, particularly at the battery terminals, are kept to a minimum. Humidity issues are especially a concern where the application is relying upon interconnects such as a battery holder for a coin cell which may not have sufficient normal force or a hermetic seal within the interconnect region. Moisture permeation into the battery packaging and subsequent reaction with the internal battery chemistry is not currently well characterized and is an area for further investigation, particularly as the allowable maximums on relative humidity has increased significantly to 85% and 90% for the ASHRAE A3 and A4 environmental standards, respectively.

ASHRAE recommends storage equipment documentation give well-defined battery replacement interval guidelines. These guidelines should be given for a range of temperature and humidity conditions. The documentation should specify whether or not high humidity levels can be tolerated. Storage equipment specifications should be clear about whether they include or exclude the batteries. Many published storage equipment non-operation specification limits are inconsistent with the environmental capability of the batteries inside the equipment.

4.4 Super-Capacitor Reliability

Another important class of charge storage devices is the so called "ultra-capacitors" or "super-capacitors" [31]. These devices are seeing growing adoption for niche applications such as replacing the battery on a RAID card. The most common type of super-capacitor is the electrical double layer capacitor (EDLC). The capacitance of a typical EDLC is in the range of 1-1000F and is several orders of magnitude higher than conventional electrolytic or film type capacitors. Unlike conventional capacitors, EDLCs use porous electrodes and no dielectric medium. They use a separator to prevent the electrodes from shorting. Due to the porosity of the electrodes, the ions in the electrolytic are in contact with a large surface area. When charged, the ions in the electrolyte are attracted to the oppositely charged electrode, thus forming a "double layer" – with two oppositely charged layers. The work (energy) done in forming the double layer (charging) can be extracted during the discharge phase. See **Figure 21** below.

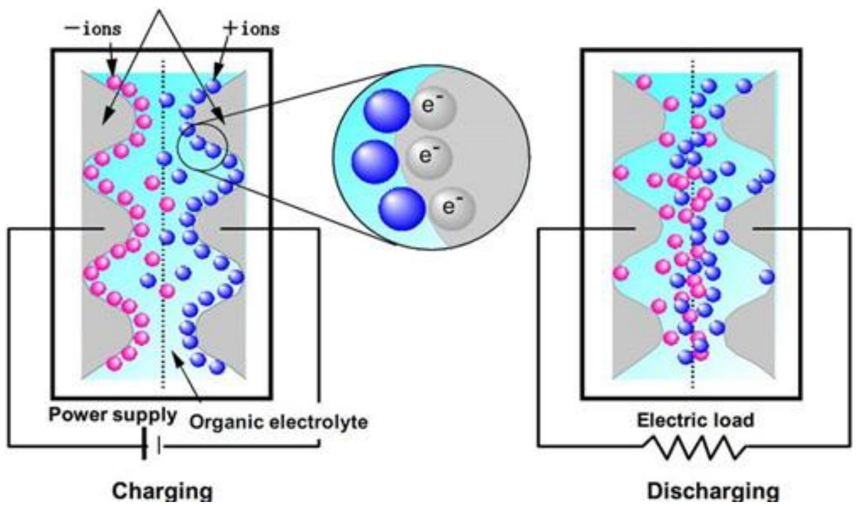

Figure 21 Charging and discharging operations for an EDLC super-capacitor [32].

A quantitative comparison of EDLC performance parameters against lead acid batteries and electrolytic capacitors is shown below in Table 11. Compared to lead acid batteries, some of the advantages of EDLCs are: a) a large number cycle times along with fast charging times, b) higher power density and faster energy release, c) smaller and lighter than batteries, d) environmentally friendly, and e) lower maintenance costs and higher reliability than batteries. Some of the disadvantages of EDLCs include: a) low operating voltages, b) shorting hazard due to low internal resistance, and c) rapidly increasing rate of degradation with temperature.

Table 11 Comparison of EDLC performance to lead acid batteries and electrolytic capacitors [31].

Performance Parameter	Lead Acid Battery	EDLC	Electrolytic Capacitor
Charge Time	1 to 5 hours	0.3 to 30 s	10^{-3} to 10^{-6} s
Discharge Time	0.3 to 3 hours	0.3 to 30 s	10^{-3} to 10^{-6} s
Energy Density (Wh/kg)	10 – 100	1 – 10	<0.1
Power Density (W/kg)	<1,000	<10,000	<100,000
Cycle Life	1,000	>500,000	>500,000
Charge/discharge efficiency	0.7 – 0.85	0.85 - 0.98	>0.95
Operating Temperature	-20 to 100°C	-40 to 70°C	-40 to 105°C

Degradation of EDLCs is accelerated primarily by exposure to high temperature and the application of over-voltage. In the case of over-voltage (see **Figure 22** below), the carbon electrode is exfoliated due to the movement of ions towards the current collector. Further increase in the applied voltage causes the electrolyte to polymerize and form a Solid Electrolyte Interphase (SEI), which degrades the performance by reducing the amount of surface area available for the formation of the double layer.

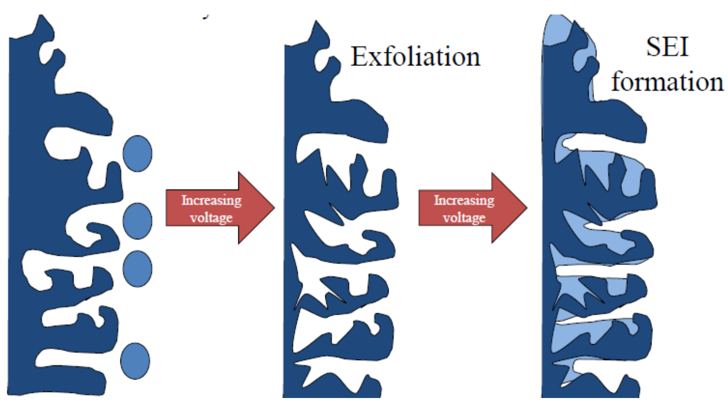

Figure 22 Performance degradation process of EDLC capacitor with the application of an over-voltage [31].

The manner in which EDLC capacitor specifications are written is similar to electrolytic capacitors – EDLCs are rated for a maximum lifetime in hours at a maximum usage temperature and a maximum voltage. The effective lifetime of an EDLC capacitor is strongly affected by the usage temperature and voltage conditions. Typical electrolytic capacitors have a rating of 1,000 hours (6 weeks) at 105C or 1,000 hours at 85C whereas a typical EDLC specification is 1,000 hours at only 70C.

Consider an example of an EDLC capacitor that is rated for 1000 hours (~6 weeks) at 70°C and 2.7V [33]. According to the manufacturer, the lifetime of the capacitor doubles for every 10°C below the maximum rated temperature of 70°C. Using the capacitor below its maximum rated voltage increases the lifetime by 1X for 2.7V (maximum rating), 4.5X for 2.25V, and 16X for 1.8V. The adjusted lifetime of the capacitor is the product of: a) the lifetime extension factor from temperature, b) the lifetime extension factor from reducing the voltage, and c) the base lifetime of the component, in this case 1,000 hours. The usage lifetime of a data center storage array can be as long as 7 to 10 years, though a 5 year refresh cycle is more typical. The practice of using a component at less than its maximum rated values (e.g. voltage, temperature) in order to increase component life and reduce failure rate is called de-rating. Because components inside the storage array chassis generate heat, the temperature of components inside the chassis is typically much higher than the inlet air temperature. As shown below in **Figure 23**, a significant amount of voltage de-rating (1.8 to 2.25V) is needed in order to extend the lifetime of a 1,000 hour 70°C rated EDLC capacitor to meet a 5 year storage array lifetime.

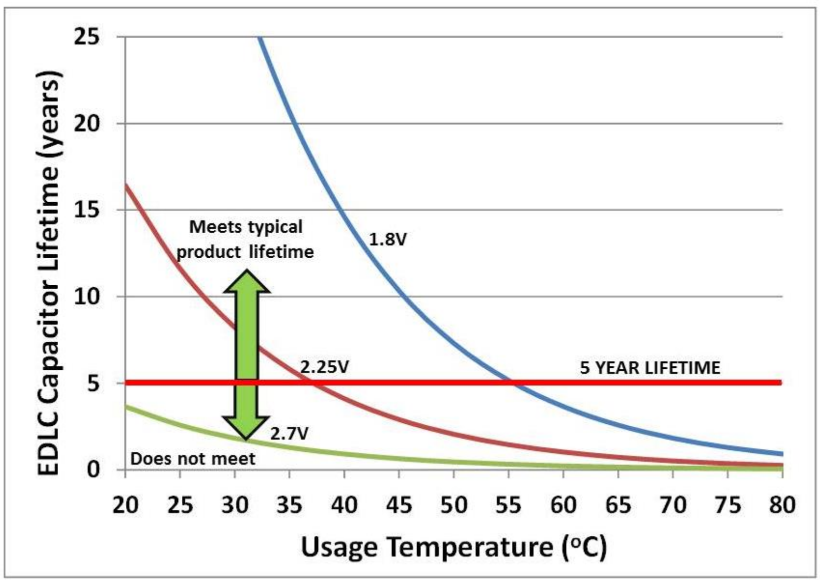

Figure 23 Example of lifetime de-rating trade-offs for an EDLC with a 1,000 hour rating at 70°C and 2.7V [33].

Another important point to consider for the application of EDLCs is the non-operational (both shipping and storage) temperature range of these components. The recommended storage temperature for some EDLCs is as low as 35°C [34]. This is well below the 45°C temperature listed in the power off conditions of ASHRAE classes A1, A2, A3 and A4. Some types of EDLC capacitors are also sensitive to high humidity levels in the range of 70%RH and above [34]. Thus, under some circumstances, it is possible for a new storage array to be shipped to a customer, see high temperatures and high humidity levels during transit, and arrive with compromised or even non-functional EDLC capacitors. Be sure to check the EDLC capacitor supplier's data sheet for storage temperature and humidity restrictions. EDLC's can be used for RAID cards and storage array applications but great care should be exercised in EDLC selection, application voltage, humidity range, and thermal design in order to make sure these components will meet the usage lifetime expectations, both operational and non-operational, of the storage equipment.

4.5 Disk Storage Array Reliability

The reliability of a disk storage array is a product of the failure rates of the disk storage media (with RAID, if applicable) as well as other components in the array such as power supplies and RAID controllers. In the 2011 Thermal Guidelines white paper [35], ASHRAE published a general failure rate for typical high volume servers as a function of air inlet temperature. In the absence of published data on the failure rate of storage arrays, this paper will assume the failure rate of storage arrays with temperature is similar to that of servers. Both servers and storage arrays use similar PCB materials, similar components, and both use disk drives. The number of disk drives in many servers is large and, in some cases, comparable to the number of drives in a storage array.

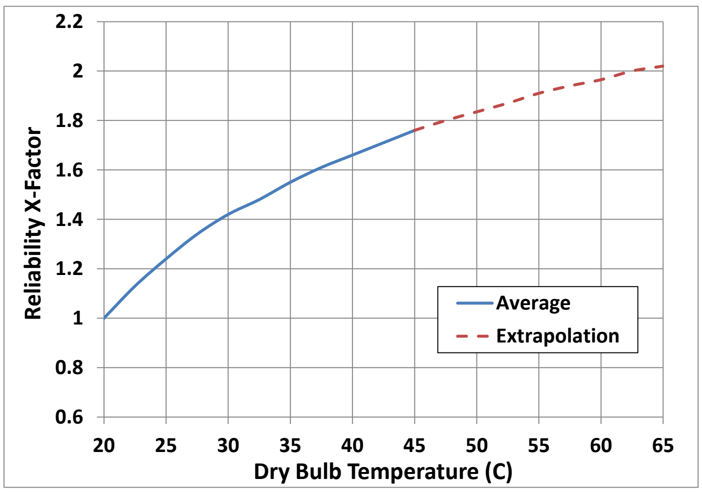

Figure 24 High volume server failure rate X-factor as a function of air inlet temperature [35].

The ASHRAE average failure rate X-factor curve above shows a failure rate increase of about 70% from 20 to 45°C. Compared to the predicted increase failure rate increase of 8.8X one would get from most MTBF prediction models such as MIL-217 and Telcordia (Arrhenius model with 0.7eV thermal activation energy), 70% is a fairly shallow increase in failure rate with temperature. The 70% value is consistent with the Google data for hard drives which also shows only a gradual increase in failure rate with temperature. In general, running disk storage arrays at cooler temperatures results in a slightly lower failure rate but higher temperature excursions won't have much of an effect on failure rate as long as they are limited to short periods of time.

As shown previously, the failure rate impact of a single drive is significantly reduced by the application of RAID. However, with increasingly higher density disks, the time required by RAID to rebuild a new disk can be hours or even as long as a day. If another disk were to fail during a rebuild window, it could cause a data loss event. To counter this problem newer RAID variants such as RAID 6 allow 2 drives to fail at the same time providing protection for the longer exposure window created by larger drives.

4.6 Tape Reliability

The success or failure of storing and recalling data depends on the reliability of both the tape drive that writes the data and the magnetic tape on which it is stored. Reliability is affected by the design of the data storage system, usage, and environmental factors. With more than 80% of the world's digital data stored on tape, tape storage is the most widely used technology for data backup and archive in long term storage environments.

Over the past 15 years, key technology improvements were borrowed from the disk drive industry that yielded much longer media life, improved drive reliability, higher duty cycles, vastly improved bit error rates and much faster data rates than previous tape technologies. This has allowed tape drive MTBF to soar from 80,000 to over 400,000 hours and media shelf life to reach 30 years [36] as shown below in **Table 12**.

Table 12 Tape drive and media reliability improvements summary.

Drive Type	Era	MTBF hours 70% duty cycle*	MTBF hours 100% duty cycle	Shelf Life Archival
3480/4480 (1st cartridge drive)	1984	35,000	24,500	<10 years
DLT 2000	1993	80,000	56,000	~10 years
SDLT 160/320/600	2002	250,000	175,000	Up to 30 years
LTO family	2000→	357,000	250,000	Up to 30 years
T9840B-D	2000→	414,000	290,000	Up to 30 years
T10000, TS11x0	2006→	414,000	290,000	Up to 30 years

*MTBF ratings are provided by tape drive manufacturers at a specified duty cycle meeting environmental specifications.

For storage devices, Bit Error Rate (BER) is quickly becoming a more popular means of measuring reliability. Current generations of tape have higher capacity and better reliability than disk. A comparison of bit error rates between different tape and hard drive technologies is shown below in **Table 13**.

Table 13 Bit error rate for different types and generations of storage technology [37].

Technology and Reliability	BER (Bit Error Rate)
Enterprise Tape (T10000x, TS11xx)	1 x 10^{19} bits
Midrange Tape LTO-5, LTO-6	1 x 10^{17} bits
Enterprise HDD (FC/SAS)	1 x 10^{16} bits
Enterprise HDD (SATA)	1 x 10^{15} bits
Desktop HDD (SATA)	1 x 10^{14} bits

For a better understanding of these reliability numbers, one can examine a study done by the National Energy Research Scientific Computing Center (NERSC). In a massive data migration exercise of over 40,000 tape cartridges, many more than 12 years old, only 35 tapes contained some data that could not be read. 99.9991% of the tapes were 100% readable. On those 35 tapes, much of the cartridge *was* read successfully. The unreadable data accounted for only 178 meters of the 22,065,763 meters of total tape [38].

Most tape storage takes place in data centers with robotic type systems. Applications that manage and secure data typically require frequent starting, stopping, and seeking by the drive. Once data has been written to tape, the tape cartridges may be stored off-site with the expectation the tapes could be re-introduced to the data center at a much later date (up to 30 years) and still reliably retrieve the data. Note that in an archive situation, only the tape cartridge is archived whereas the drive hardware (heads, electronics, motors) is not – therefore there is much less hardware to degrade over time. Less complexity results in higher reliability.

Tape cartridges have well defined environmental guidelines for long-term storage, with both temperature and relative humidity (%RH) ranges specified. These guidelines address concerns over magnetic particle oxidation, chemical changes in the particles, and the dimensional stability of the tape. An overview of how a tape drive works and the failure mechanisms that can occur in tape and tape drives is given in Appendix G.

When data is saved to tape, you want to be confident that the data will be accessible now as well as decades from now. Magnetic tape storage has one of the longest archive lifetimes, up to 30 years, of all storage solutions currently on the market. As with most materials, the physical, chemical, and magnetic properties can change as a function of environment. Minimizing these environmental changes helps ensure the robustness of data, even after long-term storage in different environments. However, in order to access the archived data one will need to carefully manage the availability of compatible tape drive hardware to read the data.

5 Flooring Design and Management

Proper floor design and management of floor loading are an important consideration, especially for disk drive based storage arrays where a single rack of equipment can easily weigh 2,000 pounds or more. The concrete slab floor of the data center must be capable of bearing the intended loads and, in some cases, the concrete slab can be a limiting factor. Types of flooring used in data centers vary but the most common is some type of raised flooring.

The design of data center flooring should comprehend the following points:

- Select access floor tiles that will meet the static and dynamic loading requirements of storage equipment. Storage equipment, especially hard disk drive based storage arrays, often requires access floor tiles with higher load ratings than most IT equipment.
- Designate a pathway for moving storage equipment to its permanent location and for routine movement of storage equipment in and out of the data center.
- Pathways will need to be made of the same access floor tile material that can handle the higher static and dynamic loads typical of storage equipment.
- If pathways are constructed from a lower load rating access floor tile material, they will need to be temporarily reinforced by laying down panels of material to spread the load such as plywood or steel plates.
- If plan is to use temporary reinforcement, data center staff must guard against someone accidentally forgetting to use the temporary reinforcement panels and damaging the pathway access floor tiles or worse causing bodily injuries.

Manufacturer floor loading guidelines should be carefully understood and followed. Racks of storage arrays with hard disk drives are often very heavy and require the use of floor tiles with a higher maximum load rating than for racks of other types of equipment. Data centers should have a heavy equipment moving procedure that all data center personnel follow. The heavy equipment moving procedure may include either the installation of a pathway of reinforced floor tiles or laying down temporary reinforcing panels such as sheets of plywood before moving heavy IT equipment. The procedure should be written by a knowledgeable facilities engineer who understands floor loading and the ratings of the floor tiles in use. Floor design should comprehend both tile loading and flexure. Exceeding load ratings can damage floor tiles and create an unsafe condition where equipment and personnel could fall through the floor. Tile flexure can indirectly create a safety hazard by causing a heavy rack of IT equipment to lean.

For more information on management of floor loading, please consult ASHRAE Datacom Series Book 5 "Structural and Vibration Guidelines for Datacom Equipment Centers" [39].

6 ASHRAE TC9.9 Recommendations

- **New ASHRAE environmental recommendations have been issued for storage array systems with hard disk drives, and for tape hardware and tape media.**
  - ✓ Many current product specifications call out a wider range of storage and non-operational conditions than the hard disk drives can support.

- ✓ ASHRAE has created separate shipping (short term) and storage (long term) recommended conditions.
- ✓ Shipping and storage are different usage conditions and need to have different recommended environmental conditions to avoid compromising any metadata or data stored on the hard disk drives.
- ✓ Industry standard conditions are issued for tape and tape library hardware for operational, non-operational, shipping, and long term storage.

Table 7 ASHRAE recommended HDD and HDD based IT equipment non-operational and operational conditions as well as operational conditions for HDDs alone.

		Temperature (^{o}C)		Humidity (%RH)[3]		Dew Point (^{o}C)	
	Time Limit	**Lower Limit**	**Upper Limit**	**Lower Limit**	**Upper Limit**	**Maximum**	**Comments**
Non-Operational - System							
Shipping[1]							
Baseline	No Limit	5^{o}C	45^{o}C	8%	90%	24^{o}C	ASHRAE Class A4
Excursion	14 days	-40^{o}C	65^{o}C	5%	95%	33^{o}C	
Long Term Storage[2]							
Allowable	Limited[4]	5^{o}C	45^{o}C	8%	90%	24^{o}C	ASHRAE Class A4
Recommended	Limited[4]	18^{o}C	27^{o}C	5.5^{o}C DP	60%	15^{o}C	ASHRAE recommended range
Operational - System							
ASHRAE Class A1	No Limit	15^{o}C	32^{o}C	20%	80%	17^{o}C	
ASHRAE Class A2	No Limit	10^{o}C	35^{o}C	20%	80%	21^{o}C	
ASHRAE Class A3	No Limit	5^{o}C	40^{o}C	8%[5]	85%	24^{o}C	
ASHRAE Class A4	No Limit	5^{o}C	45^{o}C	8%[5]	90%	24^{o}C	
Operational – Hard Disk Drive Only							
Allowable	No Limit	5^{o}C[6]	60^{o}C[6]	5%	90%	24^{o}C	

[1]Shipping assumes either: a) the system being shipped is in factory sealed shipping container and the drives have no end user data on them, or b) the hard drive is in a factory sealed plastic bag with no customer or end user data stored on it.
[2]Long term storage applies to both storage arrays as well as individual hard drives. No special packaging of the system or the drives is assumed. The drives may have end user data stored on them.
[3]Ambient air conditions must be non-condensing at all times.
[4]Hard disk drives are not intended long term or archival storage of data in a non-operational condition. For non-operational storage longer than approximately 2 years, the data should be migrated to an archival storage media such as tape.
[5]Along with a -12^{o}C minimum dew point
[6]Temperature is that of the air inside the hard disk drive chassis.

Table 8 ASHRAE recommended SSD and SSD based IT equipment non-operational and non-operational conditions as well as operational conditions for the SSD alone.

		Temperature (°C)		Humidity (%RH)[4]		Dew Point (°C)	
	Time Limit	**Lower Limit**	**Upper Limit**	**Lower Limit**	**Upper Limit**	**Maximum**	**Comments**
Non-Operational - System							
Shipping[1]							
Baseline	3 months	5°C	40°C	8%	90%	24°C	
Excursion	14 days	-40°C	65°C	5%	95%	33°C	
Long Term Storage[2]							
Allowable[3]	3 months	5°C	40°C	8%	90%	24°C	
Recommended	1 year	18°C	27°C	5.5°C DP	60%	15°C	ASHRAE recommended range
Operational - System							
ASHRAE Class A1	No Limit	15°C	32°C	20%	80%	17°C	
ASHRAE Class A2	No Limit	10°C	35°C	20%	80%	21°C	
ASHRAE Class A3	No Limit	5°C	40°C	8%[5]	85%	24°C	
ASHRAE Class A4	No Limit	5°C	45°C	8%[5]	90%	24°C	
Operational – Solid State Drive Only							
Allowable	No Limit	5°C[6]	60°C[6]	5%	90%	24°C	

[1]The SSD can tolerate higher temperatures but if the drive is shipped or stored at temperatures above 40°C then the data retention is reduced to something less than 3 months.
[2]Drive back-up should always be carried out to guard against data loss. Use of SSD media for archiving of data is not recommended.
[3]3 month data retention at 40°C per JEDEC enterprise class SSD retention use requirements [19]
[4]Ambient air conditions must be non-condensing at all times.
[5]Along with a -12°C minimum dew point
[6]Temperature is that of the SSD chassis.

Table 9 Tape Media, Tape Drive, and Tape Library Hardware Recommendations**.**

		Temperature (°C)		Humidity (%RH)*		Wet Bulb (°C)	
	Time Limit	**Lower Limit**	**Upper Limit**	**Lower Limit**	**Upper Limit**	**Max.**	**Comments**
Tape Drives & Library Hardware: Non-Operational							
Shipping Excursion	10 days	-40°C	60°C	10%	95%	29°C	Does not include tape cartridges with data
Shipping and Long Term Storage	No Limit	10°C	40°C	10%	95%	35°C	Check manufacturer spec
Tape Drives & Library Hardware: Operational							
Allowable	No Limit	10°C	40°C	10%	80%	26°C	Check manufacturer spec
Recommended	No Limit	15°C	32°C	20%	80%	26°C	
Tape Media Cartridges							
Shipping Excursion	10 days	-23°C	49°C	5%	80%	26°C	Cartridges come from the manufacturer pre-written with servo patterns
Allowable Shipping and Long Term Storage	No Limit	10°C	32°C	20%	80%	26°C	Check manufacturer specs
Recommended Long Term Storage for Archive	No Limit (up to 30 years)	15°C	25°C	15%	50%	26°C	

*Ambient air conditions must be non-condensing at all times. Per the 2011 ASHRAE Thermal Guidelines, the humidity rate of change must be <5%RH/hour for tape media, tape drives, and tape libraries.

- **Maintain IT equipment exhaust temperatures below 60°C.** Storage arrays with exhaust temperatures greater than 60°C may exceed safety touch temperature limits and may cause the air temperature in an enclosed hot aisle to exceed safe limits.
- **Protect HDDs from exposure to acoustic noise levels above 110dB sound pressure.** Acoustic noise has been shown to induce mechanical vibration that can compromise hard disk drive performance. Noise from some data center fire alarm air horns and fire suppression systems can exceeding the 110dB level can induce enough vibration to affect hard disk drive operations. Locating IT equipment farther away from the alarm horns and fire suppression systems will reduce the sound level and the corresponding induced mechanical vibration levels and this will reduce the risk to hard drive performance. It

should be noted that fans within the IT equipment can also induce mechanical vibration detrimental to hard disk drive performance. This is usually designed out by the IT equipment manufacturer.

- **Protect hard drive based IT equipment from both external vibration sources such as railroad trains, jackhammers, and heavy construction equipment in close proximity to the data center building as well as air handlers, generators and heavy equipment inside the data center.** Vibration from these sources could affect hard disk drive performance inside the systems. The site manager should be aware of strong vibration sources and inform data center personnel of these conditions. It should be noted that IT racks and cabinets usually attenuate vibration so the levels of vibration of concern are those that reach the IT equipment in the rack.
- **Exercise caution in thermal design of storage arrays that use PCIe based SSDs**. PCIe SSDs can have significantly higher power dissipation than a comparable density HDD. Temperature control systems inside IT equipment will need to respond much more quickly to temperature changes with SSDs than with HDDs.
- **Data center operators should consider implementing equipment with more sophisticated fan control algorithms from OEMs.** OEMs should make equipment with sophisticated fan algorithms more widely available. More sophisticated (vs. step) fan control algorithms respond more quickly and more accurately to work load driven temperature changes of sub-assemblies inside the storage array. This is especially important for SSDs and PCIe based SSDs which can change temperature rapidly with workload and have a very small thermal mass relative to hard disk drives. The adoption of more sophisticated fan algorithms, such as polynomial, is an energy savings opportunity for storage arrays.

Tape drives and libraries with tape media are not suitable for ASHRAE Class A3 and A4 and A4 environments. The humidity ranges of Classes A3 and A4 exceed the allowable and and recommended ranges for tape listed in

- **Table 9**. Use of tape beyond the recommended ranges may result in a higher failure rate and a higher data error rate.
- **IT equipment documentation should give battery replacement interval guidelines.** Ideally, battery replacement interval guidelines should be given for a range of data center temperatures and usage. Documentation should also specify whether or not high humidity can be tolerated. IT equipment environmental specifications should be consistent with battery environmental specifications or, if they aren't, clearly state batteries are excluded.
- **For in-rack UPS units with large batteries, be sure to consider the heat load from adjacent equipment in the rack and determine battery replacement intervals accordingly.** Battery life is a strong function of temperature.
- **Exercise caution when including super-caps in IT equipment designs.**
    - ✓ The lifetime of a supercap is very sensitive to temperature and the lifetime ratings of most supercaps are fairly short even relative to other common life-limited components such as electrolytic capacitors.
    - ✓ Supercaps may be used for IT equipment designs but only with a very careful thermal, lifetime, and de-rating analysis.
    - ✓ The IT equipment usage and storage temperature specifications should be consistent with the specifications and capability of the supercap components used. Failure to align IT equipment specifications with supercap capability could damage the supercaps resulting in data loss and corruption in the event of a power interruption.
    - ✓ One storage application where supercaps are often used is to supply power to volatile flash memory.
- **Carefully follow manufacturer floor loading guidelines for storage arrays, have a heavy equipment moving plan, and make sure data center personnel follow it.** Storage arrays, especially HDD based arrays, are very heavy and often require the installation of a higher load rated raised floor tiles than other types of IT equipment. Raised floor data centers should have a plan for moving heavy equipment in and out of the data center that includes either a reinforced pathway or temporary reinforcing panels (such as plywood) that should be in place before heavy equipment is moved. The heavy equipment moving plan should be written by a facilities engineer who understands floor loading. Failure to follow floor load ratings is a safety issue and could even result in personal injury if someone falls through the floor.
- **Design rack based storage equipment for front to back air flow.** This is needed to ensure storage equipment is compatible with energy efficient containment best practices such as hot aisle/cold aisle containment. Also, sub-systems such as controllers should not be designed with a cooling air inlet that takes in air from the side or rear of the chassis where the air may be pre-heated.
- **Make sure the temperature of the cooling air going into the front of the equipment is the same as the temperature of the cold aisle air.** Equipment installations should avoid situations where the IT equipment is taking its cooling air from enclosed spaces

inside the rack where heated air can collect and recirculate, such as an empty slot behind a blanking panel.

- **Right size the number and type of hard disk drives to your storage application for energy efficiency.** Over-provisioning your data center with high performance drives (e.g. 15k spindle speed) will increase your energy consumption. RAID levels should also be chosen for energy efficiency as RAID levels have differing power consumption values.
- **Align equipment types to separate data center environmental zones for improved data center temperature control and energy efficiency.** Although the recommended operating environmental ranges for most storage equipment fall within typical data center environments, it may be advantageous to locate some equipment in a separately controlled environmental zone for improved reliability. A zone based approach to cooling and physically locating equipment will realize power and cooling cost savings for a majority of the data center while making sure critical equipment, such as tape or batteries, is controlled within narrower specified environmental limits.
- **Avoid creating a significant back pressure on the exhaust of the IT equipment.** One example is IT equipment with a significant depth that is enclosed in a cabinet with a chimney. It may be necessary to add a "caboose" module to the back of the rack to allow more space for the air to make a right angle turn and avoid creating back pressure that could interfere with air flow and cooling.
- **ASHRAE recommends storage OEMs provide more granularity on the air flow requirements of their equipment in their customer facing documentation.** Instead of providing a single worst case value for air flow, provide a matrix or preferably a graph of typical air flow values across a typical temperature range, like the one in **Figure 16**, as a function of equipment loading and inlet air temperature. Air movement consumes a significant amount of data center power and right sizing of air flow is a significant energy savings opportunity.
- **Data center corrosion rates should be monitored using corrosion coupons on a periodic basis or with a real-time monitoring system for compliance to ISA-71 G1.** Pollution levels and corresponding corrosion rates can vary seasonally and locales that may appear to have "clean" air may, in fact, contain significant levels of gaseous pollution.
- **Protect IT equipment from particles generated by activities inside the data center such as drywall construction and drilling.** During construction temporary partitions such as plastic sheets should be hung to make sure particles don't reach the IT equipment. After any drilling or construction is finished, residual particles should be carefully wiped or vacuumed away to prevent them from being stirred up into the IT equipment cooling air stream. Gypsum particles from drywall are particularly damaging to tape heads. Fire suppression systems that use a particle based suppressant should be avoided for data centers containing tape storage systems.

7 Summary

ASHRAE TC9.9 has identified a comprehensive set of best practice recommendations for storage equipment in Data Centers. The recommendations will help Data Center operators select, install, and maintain their storage equipment for improved performance, reliability and energy efficiency. The ASHRAE recommendations also give guidance to IT equipment manufacturers on the design of their equipment and the type of customer facing documentation they need to provide. Important recommendations are also issued that will hopefully align the environmental specifications of the storage systems with those of key components inside them such as hard disk drives, super-capacitors, and batteries.

8 References

[1] F. Moore, "The Era of Colossal Content Arrives," Horison, [Online]. Available: http://www.conres.com/stuff/contentmgr/files/1/e2539a20a5546482e30833d04a894706/download/era_of_colossal_content_arrives_horison.pdf. [Accessed 7 August 2013].

[2] R. Miller, "'Digital Universe' to Add 1.8 Zettabytes in 2011," Data Center Knowledge, 28 June 2011. [Online]. Available: http://www.datacenterknowledge.com/archives/2011/06/28/digital-universe-to-add-1-8-zettabytes-in-2011/. [Accessed 7 August 2013].

[3] Gantz, John, "Extracting Value from Chaos," June 2011. [Online]. Available: http://www.emc.com/collateral/analyst-reports/idc-extracting-value-from-chaos-ar.pdf. [Accessed 7 August 2013].

[4] "Big Data Drives Big Demand for Storage, IDC Says," *Business Wire*, pp. http://www.businesswire.com/news/home/20130416005045/en/Big-Data-Drives-Big-Demand-Storage-IDC, 16 April 2013.

[5] Arcplan, "Business Intelligence Bllog," Arcplan, 23 March 2012. [Online]. Available: http://www.arcplan.com/en/blog/2012/03/big-data-faqs-a-primer/#more. [Accessed 16 January 2015].

[6] G. E. Blelloch, "Introduction to Data Compression," 31 January 2013. [Online]. Available: http://www.cs.cmu.edu/~guyb/realworld/compression.pdf. [Accessed 31 July 2014].

[7] "Data Compression," [Online]. Available: http://www.cs.princeton.edu/~rs/AlgsDS07/20Compression.pdf. [Accessed 31 July 2014].

[8] "Ultrium LTO," August 2013. [Online]. Available: http://www.lto.org/. [Accessed 8 August 2013].

[9] Datacom Equipment Power Trends and Cooling Applications, 2nd Edition, Atlanta,

Georgia: ASHRAE, ISBN/ISSN: 978-1-936504-28-2., 2012.

[10] A. Rahmanian, *Private Communication,* Dell, 2014.

[11] J. Panadero, "Characterizing Energy Efficiency in I/O System for Scientific Applications," in *ADVCOMP 2011: The Fifth International Conference on Advanced Engineering Computing and Applications in the Sciences*, Lisbon, Portugal, 2011.

[12] H. Lee, "Augmenting RAID with an SSD for Energy Relief," in *USENIX HotPower 08*, San Diego, CA, 2008.

[13] Thermal Guidelines for Data Processing Environments, 3rd edition, Atlanta, GA: ASHRAE, 2012.

[14] ASHRAE Technical Committee 9.9, Mission Critical Facilities, Technology Spaces and Electronic Equipment, Green Tips for Data Centers, Vols. ISBN: 978-1-933742-95-3, Atlanta, Georgia: ASHRAE, 2010, p. 104 pp.

[15] ASHRAE Technical Committee TC9.9, Best Practices for Datacom Facility Energy Efficiency, 2nd ed, 2nd ed., Vols. ISBN: 978-1-933742-95-3, Atlanta, Georgia: ASHRAE, 2009, p. 230 pp.

[16] Thermal Guidelines for Data Processing Environments, 3rd edition, Atlanta: ASHRAE, 2012.

[17] *NEBS Requirements: Physical Protection, GR-63-CORE Issue 4,* Piscataway, New Jersey: Telcordia Technologies, Inc., April 2012.

[18] *European Telecommunication Standard - Equipment Engineering (EE); Environmental conditions and environmental tests for telecommunications equipment, Part 1-3: Classification of environmental conditions, stationary use at weather protected..., V2.1.1,* Sophia Antipolis, France: European Telecommunications Standards Institute, 2012.

[19] JEDEC Solid State Technology Association, "JESD218A Solid-State Drive (SSD) Requirements and Endurance Test Method," JEDEC, Arlington, 2010.

[20] ASHRAE Technical Committee TC9.9, Mission Critical Facilities, Technology, Particulate and Gaseous Contamination in Datacom Environments, Second Edition, Atlanta, Georgia: ASHRAE, 2014.

[21] *ISA-71.04-2013 Environmental Conditions for Process Measurement and Control Systems: Airborne Contaminants,* Research Triangle Park, North Carolina: Instrument Society of America, 2013.

[22] F. Sun, "Feng-Bin Sun and Donald Gillis, “System Reliability with Competing Components and Load Sharing – A HDD Head-Disk Interface Perspective," *International Journal of Reliability, Quality and Safety Engineering,* vol. Vol. 18, no. No. 4, pp. pp.377-390, 2011.

[23] W. Jiang, "Are Disks the Dominator Contributor for Storage Failures? A Comprehensive Study of Storage Subsystem Failure Characteristics," *ACM Transactions on Storage,* vol. 4, no. 3, p. Article 7, November 2008.

[24] J. Elerath, "Hard-Disk Drives: The Good, the Bad, and the Ugly," *Communications of the ACM*, vol. 52, no. 6, pp. 38-45, 2009.

[25] Siemens, "Siemens Building Technologies," September 2012. [Online]. Available: http://www.buildingtechnologies.siemens.com/bt/global/en/firesafety/extinguishing/about-sinorix/latest-technical-findings/Documents/White_Paper_Potential%20damage%20to%20hard%20disk%20drives%20during%20discharges_en_September2012.pdf. [Accessed 14 February 2014].

[26] "The Availability Digest," Sombers Associates, and W.H. Highleyman, February 2011. [Online]. Available: http://www.availabilitydigest.com/public_articles/0602/inergen_noise.pdf. [Accessed 14 February 2014].

[27] B. P. Rawson, "Data Center Journal," Zimrock Marketing, 27 August 2012. [Online]. Available: http://www.datacenterjournal.com/it/inert-gas-data-center-fire-protection-and-hard-disk-drive-damage/. [Accessed 14 February 2014].

[28] e. Strom, "Effects of Humid Air on Air-Bearing Fly Height," *IEEE Transactions on Magnetics,* Vols. Vol. 43, No. 7, pp. 3301-3304, 2007.

[29] JEDEC Solid State Technology Association, "JESD218 Solid-State Drive (SSD) Requirements and Endurance Test Method," JEDEC, September 2010.

[30] T. Reddy, Linden's Handbook of Batteries, 4th Edition, New York: McGraw-Hill, 2011.

[31] M. Azarian, "Overview of Electrical Double Layer Capacitors," University of Maryland, Center for Advanced Life Cycle Engineering, College Park, Maryland, 2013.

[32] "Capacitors Blog," Suntan, 16 October 2013. [Online]. Available: http://www.capacitor.com.hk/upload/0310.jpg. [Accessed 16 October 2013].

[33] *Nichicon UM Series EDLC Capacitor Specification.*

[34] Nichicon, *Application Guidelines For Electric Double Layer Capacitors "EVerCAP" CAT.8100A.*

[35] ASHRAE Technical Committee (TC) 9.9 Mission Critical Facilities, Technology Spaces, and Electronic Equipment, "2011 Thermal Guidelines for Data Processing Environments - Expanded Data CenterClasses and Usage Guidance," ASHRAE, Atlanta Georgia, 2011.

[36] F. Moore, "F. Moore, “Tape: New Game. New Rules,” Horison, May, 2012 [Online] Available: http://www.horison.com/21stCenturyTape.pdf," may 2012. [Online]. Available: http://www.horison.com/21stCenturyTape.pdf. [Accessed 10 October 2013].

[37] F. Moore, "Horison: "Tape in the 21st Century"," April 2013. [Online]. Available: http://www.horison.com/21stCenturyTape.docx. [Accessed 16 October 2013].

[38] "National Energy Research Scientific Computing Center (NERSC), Active Archive," 2013. [Online]. Available: http://www.nersc.gov/assets/pubs_presos/AAA-Case-Study-NERSC-FINAL2-6-12.pdf. [Accessed 16 October 2013].

[39] ASHRAE Technical Committee TC9.9 Mission Critical Facilities, Technology Spaces,

and Electronic Equipment, Structural and Vibration Guidelines for Datacom Equipment Centers, Vols. ISBN/ISSN: 978-1-933742-20-5, Atlanta, Georgia: ASHRAE, 2008, p. 178 pp.

[40] Liquid Cooling Guidelines for Datacom Equipment Centers, Second Edition, Atlanta, Georgia: ASHRAE, ISBN/ISSN: 9-781-936504-67-1, 2013.

[41] M. J. Ellsworth, "An Overview of the IBM Power," *Journal of Electronic Packaging,* vol. Vol. 134, no. 020906-1, June 2012.

[42] R. C. Chu, "Review of Cooling Technologies for Computer Products," *IEEE Transactions on Device and Materials Reliability,* vol. 4, no. 4, pp. 568-585, 2004.

[43] P. E. Tuma, "The Merits of Open Bath Immresion Cooling of Datacom Equipment," in *26th IEEE SEMI-THERM Symposium,* Santa Clara, CA, 2010.

[44] T. Lee, "Fluid Selection and Property Effects in Single and Two-Phase Immersion Cooling," in *IEEE, CH2590-8/88/0000-0092*, 1988.

[45] P. E. Tuma, "Design Considerations Relating to Non-Thermal Aspects of Passive 2-Phase Immersion Cooling," in *27th IEEE SEMI-THERM Symposium,* San Jose, CA, 2011.

[46] E. Pinheiro, "Failure Trends in a Large Disk Drive Population," in *Proceedings of the 5th USENIX Conference on File and Storage Technologies (FAST '07)*, 2007.

[47] Underwriters Laboratories, "Standard for Lithium Batteries UL 1642," UL, September 2012.

[48] Underwriters Laboratories, "Standard for Household and Commercial Batteries UL 2054," UL, October 2004.

[49] (IEC), International Electrotechnical Commission, *IEC 62133-2, Secondary cells and batteries containing alkaline or other non-acid electrolytes - safety requirements for portable sealed secondary cells, and for batteries made from the, for use in portable applications,* IEC, April 2011.

[50] (IEEE), Institute of Electrical and Electronics Engineers, "IEEE Standard 1625 - 2008, IEEE Standard for Rechargeable Batteries for Multi-Cell Mobile Computing Devices," IEEE, Atlanta Georgia, October 2008.

[51] "Fuji Film," 10 September 2010. [Online]. Available: http://www.fujifilm.com/news/n100910.html. [Accessed 16 October 2013].

[52] J. Jurneke, "Applied Engineering Science: Protecting Digtal Assets for the Long Term," November 2011. [Online]. Available: http://www.fujifilm.com/news/n100910.html. [Accessed 16 October 2013].

[53] "Oracle, Sun, StorageTek: Protecting Your Archival Data With Improved Tape Dimensional Stability," January 2011. [Online]. Available: http://www.oracle.com/technetwork/articles/systems-hardware-architecture/tape-dimensional-stability-294733.pdf. [Accessed 16 October 2013].

[54] D. Floyer, "Networks Go Green," Wikibon, 7 May 2011. [Online]. Available:

http://wikibon.org/wiki/v/Networks_Go_GrEEN. [Accessed 21 October 2013].

[55] ASHRAE, Green Tips for Data Centers (Book 10 in TC9.9 Datacom Series), Atlanta Georgia: ASHRAE, 2010.

APPENDIX A – Definitions

Acronym/Term	Definition
ANSI	American National Standards Institute
AFR	Annualized Failure Rate
Archiving	Storage of data typically for the purpose of long term storage.
Backup	The copying and archiving of data so it may be used to restore the original after a data loss event.
BER	Bit Error Rate
Cold Aisle	In a data center where hot and cold air streams are contained inside the rows between racks, the cold aisle is the aisle that contains the cold air stream.
Condensing	A condition where gaseous water vapor condenses into liquid water when the temperature of the air is reduced below its dew point. Common examples of condensing conditions are: 1) a piece of IT equipment is very cold from shipping in the winter is brought into a warmer environment causing condensation on cold surfaces of the equipment, and 2) condensation occurs when a piece of IT equipment is taken from a cool air-conditioned indoor environment into a warmer and humid outdoor environment.
CPU	Central Processing Unit
DAS	Direct Attached Storage
Data Center	A building or portion of a building whose primary function is to house a computer room and its support areas; data centers typically contain high-end servers and storage products with mission-critical functions. A building or dedicated space(s) within a larger facility whose primary function is to house a computer room(s) and its support areas; data centers typically contain high end servers & storage components and related communications & connectivity hardware providing mission critical functions
Data Retention	The ability of a storage media to retain data over time and without the application of power.
Dew Point	The dew point is the temperature below which the water vapor in a volume of humid air, at a given constant barometric pressure, will condense into liquid water at the same rate at which it evaporates. Condensed water is called dew when it forms on a solid surface.
De-rating	The practice of using a component at less than its maximum

	rating (e.g. voltage, temperature) in order to prolong usable lifetime and lower the component failure rate.
Dry Bulb	The temperature of air indicated by a thermometer
ECC	Error Correction Code
EDLC	Electric Double Layer Capacitor often called a "super cap"
Equipment (IT)	Refers to, but not limited to, servers, storage products, workstations, personal computers, and transportable computers; may also be referred to as electronic equipment or IT equipment.
ESD	Electro-Static Discharge
ETSI	European Telecommunications Standards Institute
Excursion	A time limited deviation for normal or stead state conditions.
FC	Fibre Channel
FET	Field Effect Transistor
Form factor	The height of the equipment chassis measured in units of "U" where 1U = 1.75 inches
Fresh air cooling	The use of outside air for data center cooling. Another term for this is air-side economization
HBA	Host Bus Adapter
HDD	Hard Disk Drive
HDI	Head Disk Interface
Host Bus Adapter	A host controller, host adapter, or host bus adapter (HBA) connects a host system (the computer) to other network and storage devices.
Hot Aisle	In a data center where the hot and cold air streams are contained between rows of racks, this is the aisle that contains the hot exhaust air from the IT equipment
HVAC	Heating, Ventilating and Air Conditioning
IEC	International Electrotechnical Commission
Intrinsic Reliability	The reliability of the materials used to fabricate an integrated circuit or piece of IT equipment.
ISA	Instrument Society of America
ISO	International Organization for Standardization
IT	Information Technology
JBOD	Just a Bunch of Disks
JEDEC	The JEDEC Solid State Technology Association, formerly known as the Joint Electron Device Engineering Council (JEDEC), is an independent semiconductor engineering trade organization and standardization body.
LBA	Logical Bit Address

Li-Ion	Lithium Ion battery
LTO	Linear Tape Open
Metadata	Metadata is "data about data", i.e. data about containers of data. There is another type of data called descriptive metadata which is about individual instances of application data and the data content.
MLC	Multi-Level Cell flash memory
MP	Metal Particle
MTBF	Mean Time Between Failures. Mean time between failures (MTBF) is the predicted elapsed time between failures of a system during operation. MTBF is the arithmetic mean (average) time between failures of a system assuming the system is immediately repaired.
MTTF	Mean Time To Failure. MTTF measures the average time to failures with the modeling assumption that the failed system is not repaired.
MTTR	Mean Time To Repair
NAND	Negated AND or Not AND. A logic gate which produces an output that is false only if all its inputs are true; thus its output is complement to that of the AND gate.
NAS	Network Attached Storage
NEBS	Network Equipment Building Standard
Non-Condensing	An environment where the air temperature is above its dew point and there are no cool surfaces that could locally cool the air below the dew point to cause condensation.
OEM	Original Equipment Manufacturer
OR	A Boolean logic operation that is true if any of the inputs are true.
PCB	Printed Circuit Board
PCIe	PCI Express (Peripheral Component Interconnect Express), officially abbreviated as PCIe, is a high-speed serial computer expansion bus standard.
PDU	Power Distribution Unit
Rack	Frame for housing electronic equipment
RAID	Redundant Array of Inexpensive Disks
Relative humidity	a) ratio of the partial pressure or density of water vapor to the saturation pressure or density, respectively, at the same dry-bulb temperature and barometric pressure of the ambient air, b) ratio of the mole fraction of water vapor to the mole fraction of

	water vapor saturated at the same temperature and barometric pressure; at 100% relative humidity, the dry-bulb, wet-bulb, and dew-point temperatures are equal.
ROC	RAID on Chip
RPM	Revolutions Per Minute
SAN	Storage Attached Network
SAS	Serial Attached SCSI
SATA	Serial ATA
SEI	Solid Electrolyte Interphase
Server rack	A standard EIA 19 inch rack in an enclosed cabinet which is approximately 24 inches wide
SLA	Sealed Lead Acid battery
SLC	Single Level Cell flash memory
SMART	Data from a hard drive or solid state drive's self-monitoring capability that monitors temperature, usage, and errors. This data is often used to generate a warning of a possible future failure.
SSD	Solid State Drive
Storage Array Enclosure	The physical box (typically sheet metal) that holds the storage array components and sub-assemblies.
Supercap	An electric double layer capacitor (EDLC). A high-capacity electrochemical capacitor with capacitance values up to 10,000 farads at 1.2 Volts. Supercaps typically have a higher charge storage capacity than electrolytic capacitors but less than that of a battery. Supercaps are sometimes used as a power source in place of a battery when a very short duration source of power (e.g. less than a minute) is needed.
TCP/IP	TCP/IP is the internet protocol suite used by the internet and similar computer networks. TCP/IP provides end-to-end connectivity specifying how data should be formatted, addressed, transmitted, routed and received at the destination. TCP/IP is a combination of two acronyms: Transmission Control Protocol (TCP) and the Internet Protocol (IP), which were among the first networking protocols defined.
Telecom	Abbreviated term for telecommunications
TLC	Tri-Level Cell flash memory
UPS	Uninterruptable Power Supply
Volatility	Term to describe how easily electronically stored information is lost when, for example, the power supply is cut off. This term can also apply to the storage and possible loss of data

	from storage over a long period of time.
Water-side economization	The use of outdoor air to accomplish data center cooling where the means of heat transfer is water instead of air.
Wet Bulb	The wet-bulb temperature is the lowest temperature that can be reached under current ambient conditions by the evaporation of water alone. When used with the corresponding dry bulb temperature, the wet bulb temperature can be used to calculate relative humidity.
X-factor	A dimensionless metric that measures the relative hardware failure rate at a given constant equipment inlet dry-bulb temperature when compared to a baseline of the average hardware failure rate at a constant equipment inlet dry-bulb temperature of 20°C (68°F).
XOR	Exclusive OR. A Boolean logic operation that is widely used in cryptography as well as in generating parity bits for error checking and fault tolerance. XOR compares two input bits and generates one output bit. The logic is simple. If the bits are the same, the result is 0. If the bits are different, the result is 1.

APPENDIX B – Types of Storage Arrays

Network Attached Storage (NAS)

Network-attached storage (NAS) is file-level computer data storage connected to a computer network providing data access to a heterogeneous group of clients. NAS not only operates as a file server, but is specialized for this task either by its hardware, software, or configuration of those elements. NAS is often manufactured as a computer appliance – a specialized computer built from the ground up for storing and serving files – rather than simply a general purpose computer being used for the role. NAS devices can be connected via standard Ethernet networks, so they can use existing Ethernet networks. Because transactions occur at the file level, the host computer file system does not do the block-level translation, as this is handled on the NAS device; however, there is typically a higher latency associated with the TCP/IP stack. Because NAS traffic often traverses the same network as other forms of Ethernet traffic, it is more susceptible to congestion, collisions, and packet loss than other topologies.

Storage Attached Network (SAN)

A storage area network (SAN) is a dedicated network that provides access to consolidated, block level data storage. SANs are primarily used to make storage devices, such as disk arrays, tape libraries, and optical jukeboxes, accessible to servers so that the devices appear like locally attached devices to the operating system. Because the SAN is dedicated to storage traffic, it is not as susceptible to contention and congestion as a NAS and so has a much more predictable performance profile; however, it is typically more expensive to implement due to the separate infrastructure.

Direct Attach Storage (DAS)

Like the devices in a SAN, DAS are block-level devices, so the translation between the file system to logical block addressing on a target storage device must be done by the host computer. A typical DAS system is made of a data storage device (for example enclosures holding a number of hard disk drives) connected directly to a computer through a host bus adapter (HBA). Between those two points there is no network device (like hub, switch, or router). The main protocols used for DAS connections are ATA, SATA, eSATA, SCSI, SAS, and Fibre Channel. DAS devices generally tend to have the lowest latencies due to the point-to-point connections and the lack of contention in the data path; however, they do not have the connectivity options offered by a SAN device. It should be noted that NAS and SAN devices usually comprise some form of DAS devices.

APPENDIX C – Spot Cooling and Innovative Cooling Methods

There are situations in a data center where the heat generated by the storage arrays can exceed the heat removal ability of the data center. In most cases, the problem isn't a data center wide cooling problem but a localized or spot cooling issue. There are a number of commercially available solutions for spot cooling. Some of these include the rear door heat exchanger, in-row cooling, and overhead cooling solutions. Nearly all of these solutions are based on liquid as the primary heat transfer medium with an air to liquid heat exchanger in the locale where the spot cooling is needed. The details of these cooling solutions are discussed elsewhere [40] [9].

A variation of this design is described by Ellsworth et al. in [41]. In this approach an air-to-liquid heat exchanger is built into the mid bay region of the storage enclosure itself in order to cool the exhaust air from the upstream region of the chassis (see **Figure 25** below). Chilled liquid to this internal heat exchanger is supplied from a rack level Coolant Distribution Unit. The current configuration supports 384 drives, which together with the electronics, dissipate 3.5kW (2kW upstream, 1.5kW downstream) per drawer. Under nominal conditions the mid bay heat exchanger can remove 2.3kW of heat and thus chill the exhaust air from the upstream sufficiently for cooling of the electronics downstream of the heat exchanger.

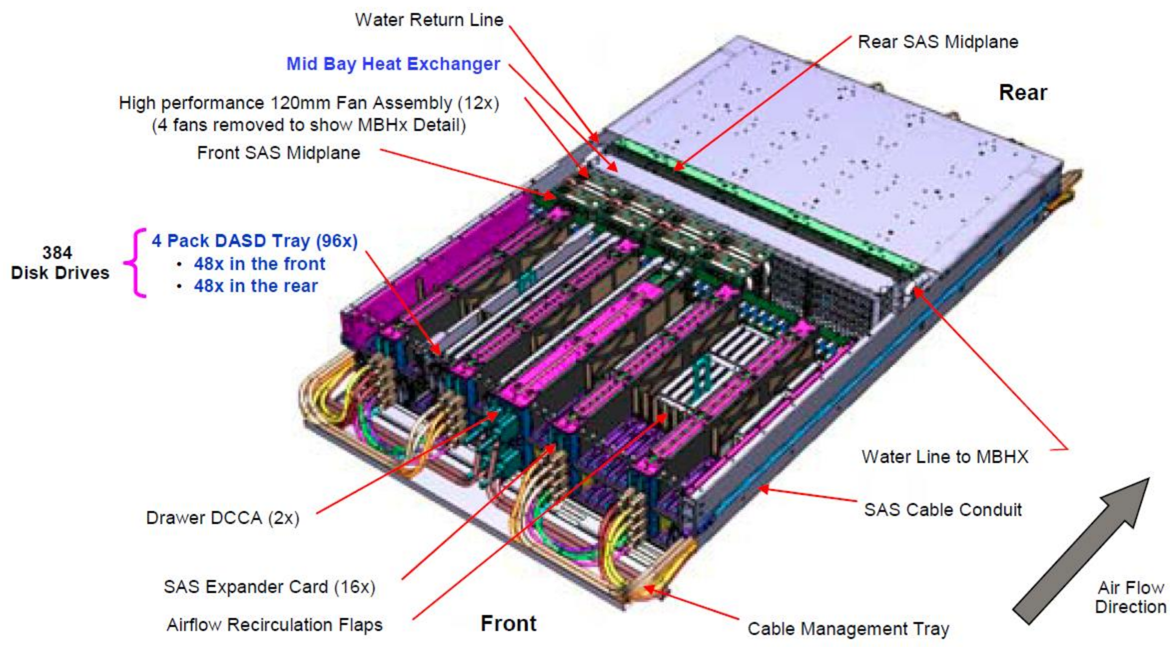

Figure 25 Example of water cooling within a disk enclosure drawer [41].

Alternatively, though not currently implemented in the industry, conduction cooling using water cooled cold plates is also a viable cooling solution. In this approach liquid cooled cold plates are designed to mate to a drive with an intermediate wet or dry interface to conductively remove heat from the surface of the drives. Direct liquid cooling is especially attractive for use with SSD cards where the higher volumetric heat dissipation, combined with a simpler package, enables simpler and more cost effective (cooling cost per watt) cooling hardware designs.

While most cooling solutions are based in part on air cooling, cooling can also be accomplished by completely immersing the IT equipment in a bath of liquid such as mineral oil (see **Figure 26** below).

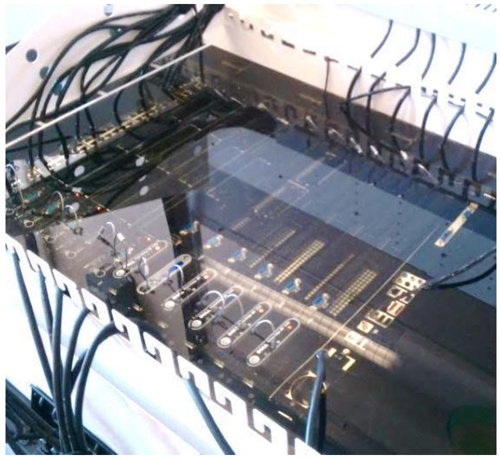

Figure 26 Photo of servers cooled by immersion in a bath of mineral oil.

Liquid immersion cooling has actually been around for many years [42] [43] [44]. A variety of liquids can be used for immersion cooling including mineral oil, Fluorinert, Novec and others. Liquid immersion cooling can be broken out into two classes: single phase and two-phase. Single phase cooling liquids, such as mineral oil, accomplish cooling solely by convective heat transfer from the IT equipment into the liquid. From there a heat exchanger removes the heat from the immersion liquid. Two phase immersion cooling liquids are engineered to have a boiling point that is below the temperature of high power components such as CPUs and FETs. The temperature of these high power components causes the liquid around those components to boil. The large latent heat of vaporization, along with the boiling mechanics, results in much higher effective heat transfer coefficients as compared to natural or forced convective cooling alone and thus significantly improved local heat removal. The disadvantage of two phase fluids is they tend to be more expensive than mineral oil and a vapor containment and condensation system is needed in order to recapture the vapor and prevent fluid loss [44] [45].

Immersion cooling requires special preparation of the IT equipment, including storage devices. Preparation includes removal of air movers such as fans and blowers from the equipment chassis. Hard disk drives require special preparation - most hard disk drives have a small vent hole that allows the pressure inside the drive to equilibrate with changes in atmospheric pressure outside the drive. For immersion cooling, this vent hole needs to be closed off to prevent the cooling fluid from entering the hard drive and interfering with the spinning media. Some companies who provide commercially available liquid cooling solutions have a process for preparing hard drives for immersion cooling, which includes hermetically sealing off the vent hole. Some hard disk drive suppliers are now offering helium filled hard drives that are already hermetically sealed. These hard drives may already be compatible with immersion cooling with no additional preparation.

An interesting point that needs investigation is whether the boiling action of 2-phase cooling fluids can cause enough vibration to interfere with HDD performance. When a liquid boils, bubbles of vapor form by cavitation. Cavitation will cause some level of vibration. Investigation is probably needed to determine whether the frequency and strength of the vibration caused by the cavitation is enough to impact disk drive performance. If vibration caused by cavitation is a problem, the IT equipment may have to rely on solid state disks instead of rotating media.

Liquid immersion cooling, in its current forms, is not feasible for use with tape media. The liquid would interfere with the mechanism of the tape drive. Another problem is the liquid cooling media tends to leach plasticizers from plastic materials making them hard and even brittle. Thus, liquid cooling isn't compatible with the plastic film used in tape media. However, as in the case of direct water cooling approaches, immersion cooling is attractive for SSD card cooling due to the compatible packaging and no additional preparatory requirements such as hermetic sealing.

APPENDIX D – Hard Disk Drive Failure

Types of operational failures that would cause an entire hard disk drive to fail include: a) bad servo track, b) can't stay on track, c) SMART limits exceeded, and d) changes in the magnetic properties of the head [24]. Servo track data is written during the manufacturing process and can't be modified in the field. Particles, contaminants, scratches and thermal asperities in the drive media can damage the servo data. SMART data from a drive is the drive's self-monitoring capability that monitors temperature, usage, and errors. This data is often used to generate a warning of a possible failure. Servo data is used to control the positioning of the read/write heads and keep them on track whether performing a read, write, or seek operation.

Another type of failure is when the head can't stay on track. During operation the head position on a hard drive is continuously measured and compared to where it should be. A loss of mechanical tolerances in the motor bearings, actuator arm bearings, noise, vibration, and servo-loop response errors can cause the head to wander and stay off track. This type of failure is sometimes intermittent and dependent on usage conditions.

Most hard disk drives have a built-in system that analyzes functional and performance data to predict impending failures called SMART (self-monitoring analysis reporting technology). SMART monitors the number and frequency of sector re-allocations as well as the number of remaining spare sectors on the HDD. If an excessive number of re-allocations occur within a defined time interval, the SMART algorithm will deem the drive as unreliable and will deem it a failing drive.

Failure of an entire disk drive can also happen when there is a change in the magnetic properties of the head. ESD events, high temperatures and physical impacts from particulate contaminants inside the drive can cause the magnetic properties of the head to change causing a drive failure. Drive failures can also be caused by component failures on the PCB such as a failing DRAM or cracked chip capacitor [24].

Hard disk drive failure rates can vary widely between manufacturers, models and manufacturing vintage. Factors responsible for this variation include: a) the amount of airborne contamination (particles) inside the drive, b) design changes – while most of these are beneficial to performance, some degrade reliability, and c) changes to the drive manufacturing process to improve yield – some of yield improvements may degrade reliability.

One of the most comprehensive hard disk drive reliability studies was carried out by Google [46]. The Google team analyzed failure data from a very large disk drive population deployed across multiple data centers over a long period of time. They analyzed the data on the basis of the age of the drives, drive utilization and drive temperature. Several of the key figures from the Google paper are shown below. It should be noted that the value of reliability index, such as MTTF and/or AFR, is dependent on the failure criteria. The failure definition in Google paper is "a drive is considered to have failed if it was replaced as part of a repairs procedure". Therefore, a more appropriate term for AFR in Google paper should be ARR, the annual replacement rate. Since each HDD manufacturer may have its own failure criteria, the MTTF and AFR numbers may differ.

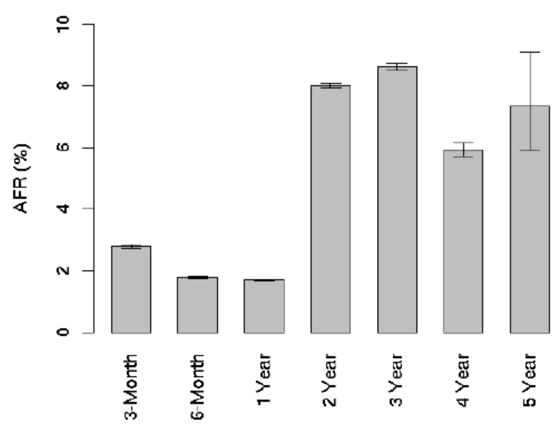

Figure 27 Annual Failure Rate percentage (AFR) as a function of disk drive age [46].

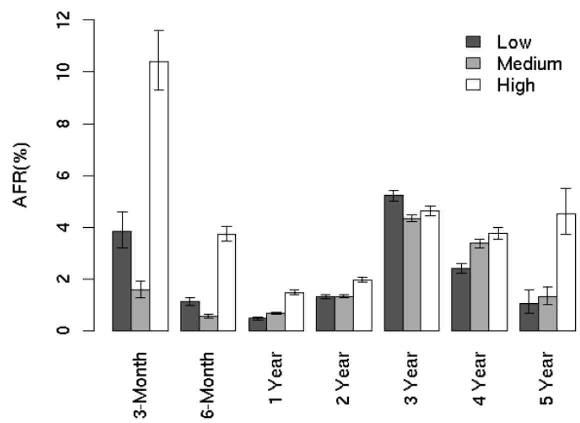

Figure 28 Annual Failure Rate percentage (AFR) as a function of disk drive utilization and age [46].

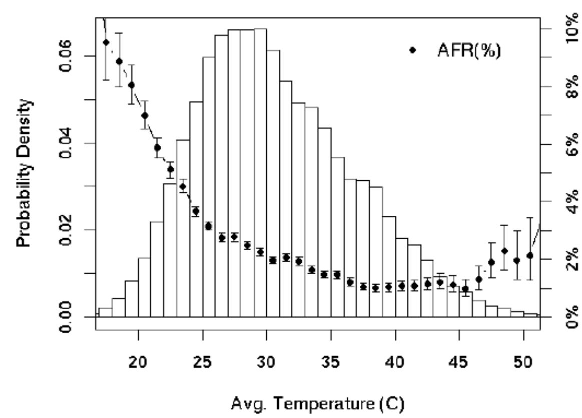

Figure 29 Probability density and Annualized Failure Rate (AFR) percentage as a function of disk drive temperature [46].

Results of the Google study show; a) noticeable influence of infant mortality phenomena during first year of field operation, b) failure rate increasing with drive age but only after 2 years, c) slightly higher failure rates associated with high drive utilization, and d) highest failure rates at low temperature with only a slight increase in failure rate above 45°C. Of all of the findings the temperature is probably the most surprising – it contradicts the traditional assumption of failure rates doubling for every 10°C of temperature rise.

APPENDIX E – Flash Media Reliability and Management

A diagram of a simple NAND flash memory cell is shown below.

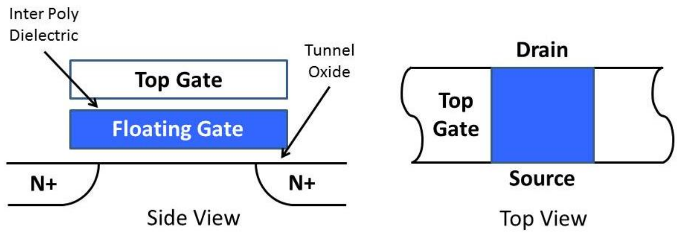

Figure 30 Cross-section and top views of NAND flash memory cell.

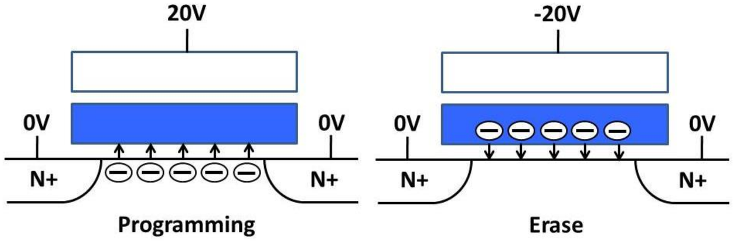

Figure 31 Illustration of programming and erase operations on flash memory cell.

Flash memory stores a bit of information by moving charge onto or off of a floating gate. One of the advantages flash memory has over DRAM is the information in the flash memory cell can be stored for extended periods of time without the need to have external power continuously applied in order to retain the data. There are two types of flash memory: SLC and MLC. SLC is single level cell and stores only a single bit of information in each memory cell. MLC is multiple level cell and can store two or even three bits (TLC) of information in a given memory cell. MLC is advantageous in that it affords higher density drives and lower cost. Both types of flash memory have a finite number of write-erase cycles they can tolerate before they begin to fail. However, MLC flash is more susceptible to degradation and fails at a shorter number of write-erase cycles than does SLC flash. An illustration of how MLC flash works and how degradation affects it is shown below.

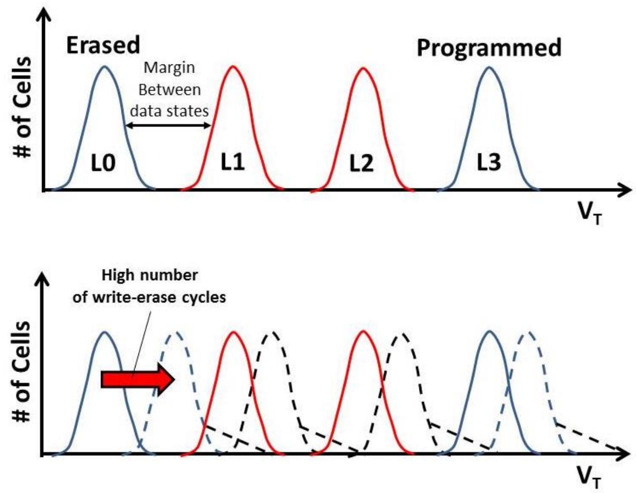

Figure 32 MLC voltage level distributions for storing two bits (top graph). Illustration of how wear out compromises margin between data states (bottom graph).

The most common type of MLC flash uses four voltage levels to store two bits of data per cell as shown in **Figure 32**. As those cells are degraded by repeated write-erase cycles and associated wear out mechanisms the shape of the voltage distributions changes and shifts to higher voltages. This erodes the margin between data states, i.e. between the voltage the SSD is expecting and the actual voltage state of the cell. Wear out is reached when the margin between voltage levels is so small that different data states can no longer be consistently resolved from one another with a high level of confidence.

Solid state disk drives employ a number of strategies to manage and mitigate the wear-out behavior of the flash memory. These include:

- Wear leveling
- Error correction code (ECC)
- Bad block retirement

One means of extending the life of a drive is to make sure the number of write-erase cycles are fairly uniform across the array. That is, to spread the wear uniformly across the memory array so no one physical location is subject to a disproportionate amount of wear. This technique is called wear leveling. It is accomplished by dynamically changing electrical to physical map of where data is stored on the drive to prevent any set of physical locations from receiving a disproportionately high amount of write-erase cycles.

Error correction codes (ECC), also called hamming codes, were originally developed to protect data from spacecraft where the chance of a cosmic ray single event upset was high. The most basic type of ECC is implemented by adding enough extra parity bits to a page so the original data can be reconstructed when a certain number of the original bits have been corrupted. The error correction codes used in SSDs are a combination of very advanced ECC schemes. Error correction codes aren't free – they consume additional memory and they can place a small tax on performance so the SSD manufacturers implement them to the extent needed by the type of media they are using (SLC vs. MLC) and to meet the bit error rate expectations of the end use market. ECC schemes will likely increase the latency of the SSD over its lifetime.

Another means of managing the flash media for higher reliability is bad block retirement. When a flash memory component identifies a bad block containing bits that won't program or erase, it will flag that physical block of memory as bad. The solid state drive itself will then manage the bad block by moving the data to a different block and masking the bad block from its look-up table of available physical locations.

APPENDIX F – Summary of Battery Technologies

Table 14 Battery technology and characteristics [47], [48], [49], [50].

	Battery Chemistry	Where Used	Advantages	Disadvantages
Rechargeable	Lithium Ion (Li-Ion) Lithium Polymer Lithium Iron Phosphate	✓RAID Adapter Cards ✓Storage Battery Backups ✓OEM products	✓High Energy Density ✓Low maintenance ✓Low Self Discharge ✓Lightweight ✓Cycle Life is Good	✓Explosive Over Charge – Protection circuit per cell ✓Cost ✓Unknown performance in standby applications
Rechargeable	Nickel Metal Hydride (NiMH)	✓RAID Adapter Cards ✓OEM products	✓Good Energy Density ✓Simple storage and transportation ✓Environmentally friendly	✓High Self Discharge Rate ✓Complex charge algorithm ✓Cost more than SLA ✓Fading technology for portable applications
Rechargeable	Sealed Lead Acid (SLA)	✓Large System Battery Backups ✓UPS'es	✓Inexpensive ✓Low self discharge ✓Capable of high currents ✓Common form factors	✓Lead = Environmental Concern ✓Low Energy Density – Bulky ✓Must be stored charged ✓Limited Cycle Life
Non	Primary Batteries ($LiMnO_2$, LiC_nF)	✓Card assemblies for powering small loads (e.g., a Time Of Day clock)	✓High Capacity per unit volume ✓Long shelf life ✓Cost ✓Familiar technology	✓Non rechargeable ✓Best for low rate discharges

Increasing Energy Density / Increasing Cost

Table 15 Lithium ion battery chemistry comparison.

Battery Chemistry	Where Used	Advantages	Disadvantages
Lithium Iron Phosphate	✓Storage Battery Backups ✓OEM products	✓Safest of Li-ion technologies. ✓Low capacity loss over time at high temperature. ✓Capable of high discharge rate due to low cell impedance ✓Long life.	✓Lower energy density than other Li-ion technologies. ✓Flat discharge curve increases gauging difficulty.
Lithium Cobalt	✓Limited Uses	✓Highest energy density of current Li-ion chemistries.	✓High reactivity results in safety concerns ✓Limited cell types available. ✓Currently limited development.
Lithium Nickel Manganese Cobalt	✓RAID Adapter Cards ✓Storage Battery Backups ✓OEM products	✓High energy density.	✓Trade-off between capacity and discharge rate capability.
Lithium Polymer	✓RAID Adapter Cards ✓Storage Battery Backups ✓OEM products	✓High space efficiency.	✓Trade-off between capacity and discharge rate capability. ✓Over-charge or over-discharge can cause gassing leading to bulging of battery packs.

Table 16 Data center battery technology comparison.

Battery Chemistry	Where Used	Advantages	Disadvantages
Flooded Lead Acid	✓Most North American Data Centers	✓Longer life than SLA ✓Low self discharge ✓Capable of high currents ✓Common form factors	✓Ongoing preventative maintenance required ✓Higher initial cost than SLA ✓Lead = Environmental Concern ✓Low Energy Density – Bulky ✓Must be stored charged
Sealed Lead Acid (SLA)	✓Some Data Centers in Asia, Latin America, and Europe	✓Inexpensive (initial cost) ✓Low maintenance ✓Low self discharge ✓Capable of high currents ✓Common form factors	✓Less life than Flooded Lead Acid ✓Lead = Environmental Concern ✓Low Energy Density – Bulky ✓Must be stored charged ✓Limited Cycle Life
Lithium Ion (Li-Ion)	✓Primarily test case sites	✓Smaller size due to higher energy density ✓Low maintenance ✓Relatively good cycle life performance	✓Cost ✓Safety concerns ✓More complex electronics ✓Life expectations in standby applications not well understood yet

APPENDIX G – Tape Mechanics and Failure Mechanisms

This section will outline the potential failure mechanisms in tape storage and how tape drive design and magnetic tape are able to offer excellent reliable data availability to customers. Most tape storage occurs in data centers with heavy data retrieval in robotic environments. Applications that manage and secure data typically require frequent starting, stopping, and seeking by the drive. Once data has been written to tape, the tape cartridges may be stored off-site with the expectation the tapes could be re-introduced to the data center at a much later date (up to 30 years) and still reliably retrieve the data. Note that in an archive situation, only the tape cartridge is archived whereas the drive hardware (heads, electronics, motors) is not – much less hardware to degrade over time. Less complexity results in higher reliability.

A brief overview of how a tape drive works can help identify potential failure mechanisms. At the highest level, there are 2 components to a tape drive system: the tape drive and the tape cartridge.

Within the tape drive, the types of possible operational failures are very similar to those previously described for hard disk drives: bad servo track, can't stay on track (lateral tape motion), and changes in the magnetic properties of the recording head.

A key difference between disk drives and tape drives is the number of data tracks that are simultaneously written or read. Instead of relying on a single track for writing and reading the disk, tape drives utilize heads with between 8 and 32 tracks. With this multiple track format, data can be distributed across many tracks. Loss of several writers or readers on the recording head does not translate into loss of data with the redundancy designed into tape recording formats. The more channels used to put data on tape, the higher the number of simultaneous errors may be corrected by the error correction code (ECC). This creates a significant improvement over other ECC systems.

Tape does not have to manage defects. With the ability to read the data at time of recording (read verified writes), the tape drive knows if the data is intact. Should the data be found either bad or marginal, tape drives will automatically skip across sections of media if it encounters difficulty in writing, remapping the data farther down the tape rather than trying to rewrite again in the same location.

Turning to the tape itself where the data is stored, the failure mechanisms can be classified as Mechanical, Magnetic, or Chemical as show below in

Table 17.

Table 17 Summary of Tape failure mechanisms.

Failure Mechanism	Description	Impact	Level of Risk
Chemical	Corrosion of magnetic particles	Potential data loss; high error rate	Medium. Reduced with the latest generation of BaFe magnetic particles [51]
	Lubricant/binder breakdown	Can lead to durability and runability issues; potential data loss; high error rate	High. Store in cooler, dry environments. Avoid extended exposure to high temperatures (>60°C)
Magnetic	'Bit rot' - demagnetization of magnetic particles	Data reliability issue if more than 10% demagnetization loss	Low. Data suggests risk is extremely low [51]
Mechanical	Edge damage	Folds or creases at edge of tape can lead to loss of data read back reliability due to lateral tape motion/off-tracks	Medium. Even if cartridge is dropped, majority of tape data is recoverable [52]
	Durability/Wear	Magnetic coating may wear off. This leads to higher error rates.	High. Migrate data off of tape cartridges nearing manufacturers' wear-out limit - based on total meters of tape cycled.
	Tape Dimensional Stability	Change in tape dimensions can lead to mis-registration between read head and data track (read back reliability)	High. Minimize changes in temperature and humidity over storage life of tape [53]

Folds or creases at the edge of tape can lead to loss of data read back reliability due to lateral tape motion and the readers going off-track. This sort of tape damage can occur through excessive use within the drive tape path or through improper handling (dropping the tape cartridge). To mitigate these mechanical damage risks, tape drive design focuses on tape speed, design of the tape path and guiding, as well as ruggedizing the cartridge.

Tape cartridges will not last forever running in tape drives. Since the tape, and the data on it, is in direct contact with the recording heads, the magnetic coating may begin to wear away leading to higher error rates after extended use. For many archive data applications, this is not a concern. Archive storage typically will write data to the cartridge, filling it to capacity, and then store it with only occasional read access. For customers who utilize tape for backup, re-using the same cartridges multiple times, it is recommended that data be migrated off of tape cartridges when the cartridges near the manufacturers' wear-out limit. The wear out limit is based on the total number of meters of tape cycled across the recording head and is typically tens of millions of meters. Cartridges also have a limited number of loads, 10,000 to 25,000, before mechanical failure.

Tape cartridges have environmental guidelines for long-term storage, with both temperature and relative humidity (%RH) ranges specified. Often, these guidelines are assumed to address concerns over magnetic particle oxidation or chemical changes. However, another purpose of the environmental storage guidelines is to address the dimensional stability of the tape. As with all materials, magnetic tape can change its dimensions when exposed to different temperatures and humidity ranges. One example, on the macroscopic level, would be how a wooden door swells in high humidity making it difficult to open. In many cases, these dimensional changes are not noticeable. However, data bits on magnetic tape are tiny enough that small changes in the dimensions of the tape can impact the likelihood of successfully recalling data.

To understand this better, examine **Figure 33.** Data is initially written under recommended environmental conditions with excellent read back quality. Data is written on the tape with multiple writers on the tape head. The distance between each written track is controlled by the spacing of the writers on the recording head. The magnetic data bits on the tape are well aligned with the magnetic recording head. The tape is then stored, under tension, at a certain temperature and humidity. Over time, the width of the tape may shrink or expand. Physical properties of the tape, such as the coefficients of temperature, humidity, and tension narrowing, determine the amount of dimensional change. Finally, when the tape cartridge is re-read, the environmental conditions may be different. The spacing between the data tracks written on the tape will have changed. The data written on the tape now is shifted in relation to its original location. If there is too great a difference between the spacing of the written tracks on tape and the readers on the recording head, the data cannot be read back.

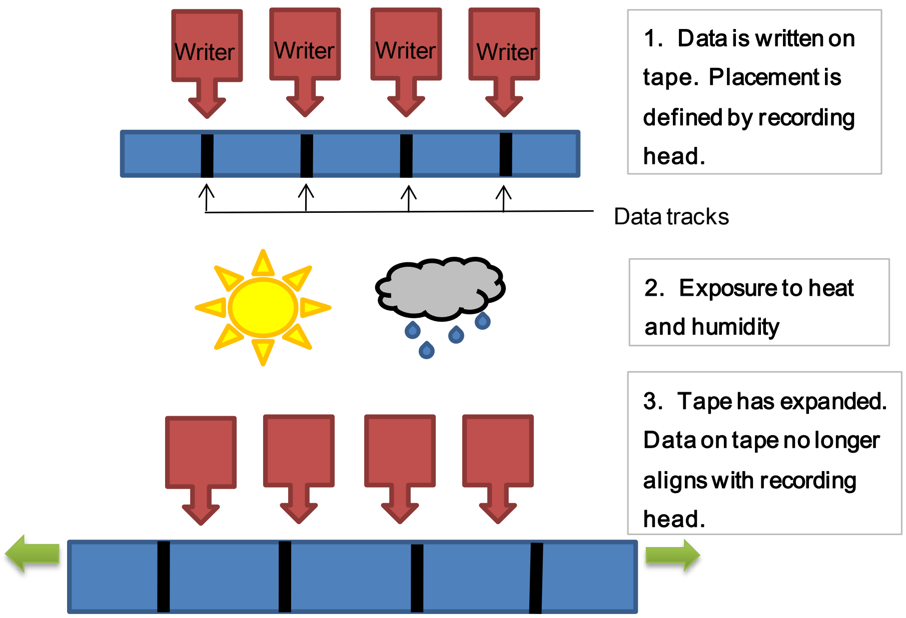

Figure 33 Small changes in the dimensions of the tape can affect the read back quality of data **[53]**.

With data stored in magnetic bits on the tape, demagnetization of the magnetic particles, attractively known as 'bit rot,' can be a concern. Data reliability is an issue if more than 10% demagnetization loss occurs. In accelerated testing to simulate 30 year storage, Fujifilm has shown that MP (metal particle) tapes have up to an 8% loss and Barium Ferrite (BaFe) tapes have 0% loss [51]. These losses are not enough to affect the stability of the magnetic properties under long term storage conditions.

The last magnetic tape failure mechanism is chemical. If the magnetic particles undergo a chemical reaction or corrosion, there is potential for high error rates and data loss. Obviously, environmental conditions and contaminants in the air play a role. Current magnetic particles either have a passivation layer (metal particle tape) or are chemically stable (non-reactive, BaFe tape) to mitigate the potential for chemical induced failure. The binder and lubricant used in the tape also can undergo chemical reaction or breakdown. Changes in the lubricant and/or binder can lead to tape durability issues with potential for high error rates and data loss. The best mitigation to this risk is to store tapes in cooler, dry environments. Avoiding extended exposure to temperatures greater than 60°C is strongly recommended.

APPENDIX H – Raised Flooring Systems and Related Industry Standards

A typical raised flooring system consists of a framework of vertical and horizontal supports with tiles placed on top (see **Figure 34** below).

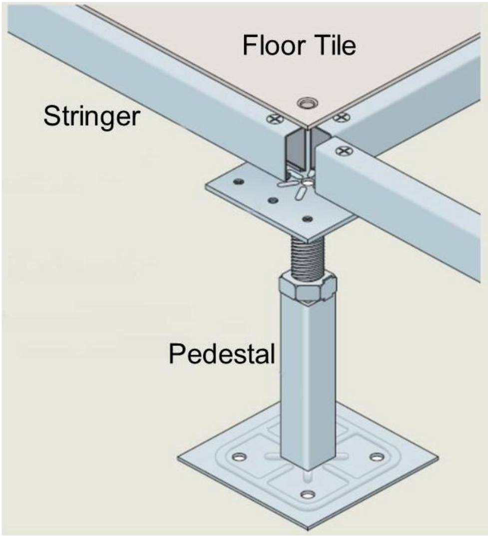

Figure 34 Example of a raised flooring system with vertical supports (pedestals) and horizontal supports (stringers).

There are several different types of load ratings for access floor tiles: static loads and dynamic loads. Static load is the rating of the access floor tile to support a stationary load. This load can be either concentrated or uniform. Uniform loading is when the load is uniformly distributed over the entire surface of the floor panel. However, the most common type stationary load on a floor tile occurs when the load is concentrated on just a few points such as the four wheels of an equipment rack. The size of equipment racks relative to the size of the floor tiles is such that a single floor tile may be supporting only one or two feet of the equipment rack. At first glance it would appear one could calculate the static load on a floor tile by taking the weight of the rack, dividing it by the number of feet or wheels supporting the rack, and then multiplying that result by the number of wheels or feet on the particular floor tile. However, the weight distribution of a typical IT equipment rack is more complex. IT equipment racks tend to have the equipment mounted to the front rails of the rack with the front feet or wheels of the rack bearing a disproportionate amount of the load. When doing floor tile loading calculations, one should take into account a potentially non-uniform weight distribution of the rack and do the calculations for the worst case scenario where the feet or wheels resting on a particular floor tile are the heaviest loaded wheels. This is true for both static load and dynamic or rolling load calculations. Uniform loading assumption may not always be true and may mislead a floor design to believe the tiles are capable of supporting the load when, in fact, they may not be.

The most common dynamic load occurs when a rack is being rolled across the raised floor. Maximum rated rolling loads for raised floors are typically lower than maximum static load ratings. Loads that exceed maximum ratings can permanently deform the access floor tiles creating an unsafe condition and necessitating replacement of the floor tiles. Examples of static and dynamic loads are shown below in **Figure 35**.

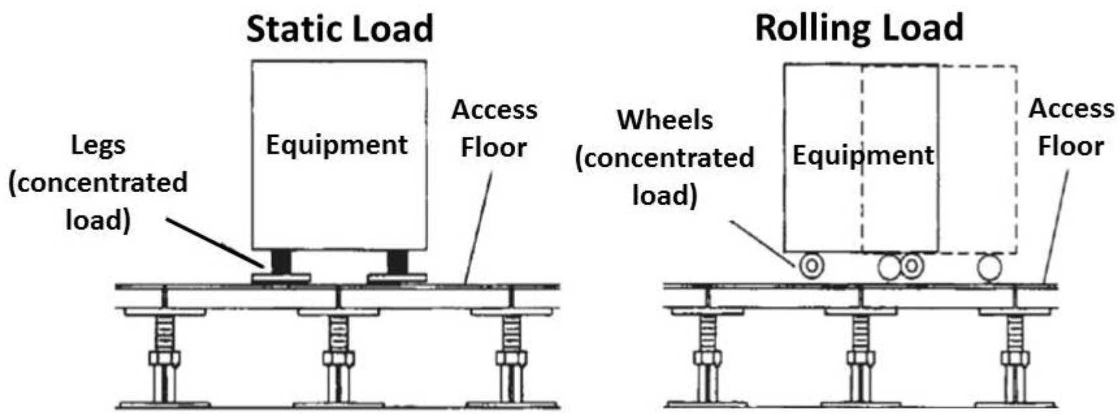

Figure 35 Example of static concentrated load and a rolling load on a raised access floor.

Some of the most common flooring specifications and testing recommendations are listed below.

Table 18 Common access floor tile specifications and recommended testing methods.

Title	Spec Number	Country
CISCA Recommended Test Procedures for Access Floors - 2007	NA	USA
Property Services Agency (PSA) Method of Building Performance Specification Platform Floors (Raised Access Floors)	MOB PF2 PS	UK
Raised access floors	EN 12825:2001/BS EN 12825:2001	Europe/UK
Telecommunications Infrastructure Standard for Data Centers	ANSI/TIA-942-2005	USA