Reliability Analysis of Disk Drive Failure Mechanisms

Sandeep Shah, Network Appliance
Jon G. Elerath, Network Appliance

Key Words: Disk drive, Reliability, Hazard rate, Failure Mechanism, Error messages, Field data

SUMMARY & CONCLUSIONS

Reliability analyses are performed on the field failure data of various drive families from different drive manufacturers to gain insight into the nature of the underlying failure mechanisms and their contribution to the overall failure rate of the disk drive. The effect of various disk drive firmware changes implemented throughout the operating life of the disk drive and the capacity dependent failure mechanisms are studied in detail.

The results obtained from the analysis of three disk drive families shows that a top-level failure rate analysis is not adequate to understand and improve the reliability of disk drives. Detailed hazard rate analyses of all failure mechanisms should be performed to understand the top contributors to the overall hazard rate. The results obtained from the analysis of one drive family shows that the composite hazard rate obtained from the combination of various failure mechanisms depends on which failure mechanism is dominating at what point during the operating life. On this particular drive, the authors conclude that in early stages of the life of the drive, one failure mechanism is dominating while a second failure mechanism starts dominating towards the end. The rest of the failure mechanisms are either decreasing or constant and have a minimal effect on the overall hazard rate. Analyses of a second drive family show that a firmware change implemented on the drive to fix a particular problem in the field inadvertently accelerated some hardware failure mechanisms resulting in the increase in overall hazard rate of drive. The third drive family analyses show that the capacity dependent failure mechanisms on the higher capacity disk drive contributed significantly to the increase in the overall hazard rate.

1. *INTRODUCTION*

Disk drives are complex electromechanical systems. Additionally, sophisticated firmware and software are required for the drive to work with other components in the system. Due to this complexity, it is sometimes very difficult to understand and improve the reliability of disk drives and their components. Only after a detailed failure analysis is completed, the failure mechanisms fully understood, and the nature of the failure rate understood, whether increasing, decreasing or constant, can corrective actions be implemented and drive reliability improved. Each failure mechanism contributes to the overall failure rate of the disk drive in a different fashion at different times in the operating life of the disk drive. For example, one failure mechanism may have a high but constant failure rate, while another failure mechanism may have a low but increasing rate. It is not possible to fix all failure mechanisms, so it is essential to understand and fix the key failure mechanisms contributing most to the overall failure rate of disk drives. In general, infant mortality failures include firmware failures, particulate contamination failures, media defects, manufacturing process induced or handling type failures. Hardware failures, such as electrical component failures and head/media interface failures dominate the useful life of the disk drive. Long-term degradation is due to motor failures, contamination or corrosion.

In the disk drive industry, it is difficult to determine the underlying failure mechanisms from the data obtained from the field. So, often only drive level metrics are reported. One such metric to express reliability at the drive level is Annualized Failure Rate (AFR), which is shown in Figure 1. This is a three month rolling average that does not account for failure rates which change as a function of time or vintage. The AFR assumes, somewhat implicitly, a constant failure rate for all drives and that probability of failure is equally likely in any fixed period of time. This means the probability of failure in the first 100 hours of operation is the same as the 100 hours between 43,800 and 43,900 hours of operation. NetApp Reliability Engineering has published studies of field data that show this is rarely true (Ref. 1).

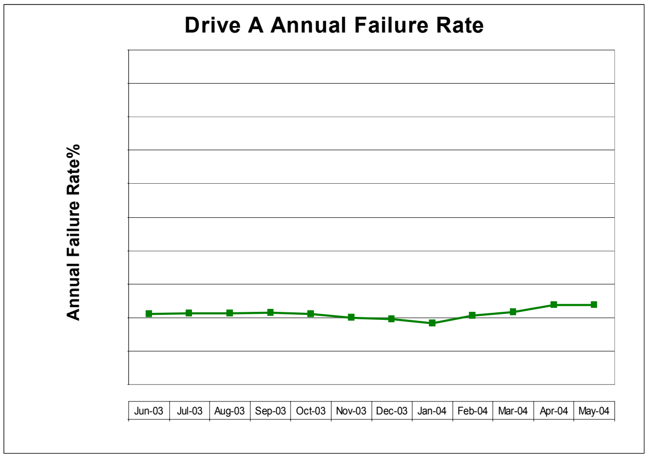

Figure 1. Annual Failure Rate of Drive A

To account for the changing failure rate nature of disk drives, a more sophisticated time-dependent analysis like

Weibull analysis or hazard rate analysis is performed. Figure 2 is a linear plot of failure rate as a function of time of drive family A. While this type of analysis addresses the changing failure rate problem, it does not give insight into the top issues affecting the overall reliability of disk drive. Hence a hazard rate analysis based on failure mechanism needs to be done to get better visibility into the underlying failure mechanisms and their contribution to the overall hazard rate.

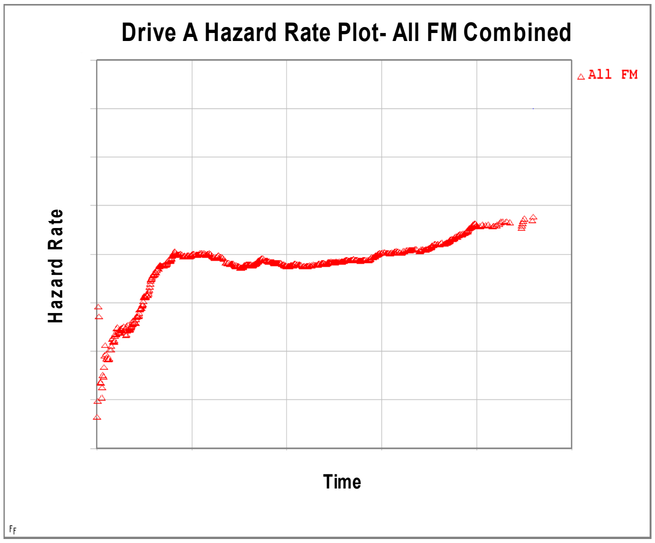

Figure 2. Hazard Rate Plot of All Failure Mechanisms Combined of Drive A

2. *DATA SOURCES*

2.1 *Field Returns and Failure Analyses*

Network Appliance incorporates multiple disk drive families from the world's leading manufacturers into its systems. There is a large field installation base and an extensive reliability database (thousands of drives for each drive family). For every drive shipped, this highly sophisticated database tracks the serial number, the date it shipped, the date it is suspected as a failure and returned by the customer, and, to some extent, the reason for the failure (error code). After being received back from the customer, each failed drive undergoes a first level failure analysis and some undergo detailed failure analysis. These data allow time-dependent analyses on different error codes or failure mechanisms with high accuracy. The data are analyzed for fit to Weibull and Lognormal distributions using the Maximum Likelihood Estimates (MLE) method and goodness of fit tests. Based on the best fit of the data, a liner plot is made using the Super Smith (Ref. 2) software. To maintain manufacturers anonymity, the drive families have been identified as A, B and C.

2.2 *Support Logs*

A large portion of Network Appliance systems record events that occur during operation. Most log entries are records of benign events or information. However, the logs also record the drive serial number and the specific disk drive performance and behavior characteristics leading up to a drive event or failure. Using these logs and the recorded error codes, events can be mapped to the failure symptoms and modes through the drive serial number. After failure analysis, the symptoms and modes are mapped to the failure mechanisms. These logs are critical to the complete failure analysis and corrective action process as it stores vital information about the drives health before failure.

3. *FAILURE MODES AND MECHANISMS*

Error codes are fairly synonymous with failure mode. However, there is not a one-to-one mapping of mechanisms to error codes (or failure modes). A description of some of the error codes is provided in the next section, along with common failure mechanisms.

3.1 *Failure Modes*

Disk drives have many different failure modes. One common structure for reporting errors is defined by the Small Computer Systems Interface (SCSI) standard. The standard prescribes the structure of the data and the command set for its retrieval. The structure is to return 3 pairs of hexadecimal values. These pairs are (in order) the Sense Key, Sense Code and Sense Code Qualifier. While the specific meaning of the Sense Keys and Sense Codes are fairly consistent across drive manufacturers, the Sense Code Qualifier varies by manufacturer. The meanings of the Sense Key are in Table 1. The quantity of Additional Sense Codes and Additional Sense Code Qualifiers is too great to present in this paper, but a few are shown in Table 2 as examples of the types of information they provide. For example, Sense Key "01" means Recovered Error. The Additional Sense Code "18" means recovered with error correcting codes (ECC), and an Additional Sense Code Qualifier of "01" means retries were employed to recover the data. Some disk drive manufacturers provide an additional set of hexadecimal values called Sense FRU Code, which provides one more layer of information on the failure mode.

SCSI Sense Key	Description
01	Recovered Error
02	Not ready
03	Medium error
04	Hardware error
05	Illegal request
06	Unit attention
07	Data protect
08	Blank check
09	Firmware error
0B	Aborted command
0C	Equal
0D	Volume overflow
0E	Miscompare

Table 1. Sense Key Codes and Descriptions

01 Recovered Error	Description
01-17-01	Recovered Data with Retries
01-17-02	Recovered Data using positive offset
01-17-03	Recovered Data using negative offset
01-18-00	Recovered Data with ECC (no retry attempted)
01-18-01	Recovered Data with ECC and retries applied
01-18-02	Recovered Data with ECC and/or retries, data auto reallocated
01-18-06	Recovered Data with ECC and offsets
01-18-07	Recovered Data with ECC and data rewritten

Table 2. Example Error Codes and Descriptions

nature of their designs, specific drive families from specific manufacturers tend to have their own set of mechanisms that produce the majority of errors for each code. For example, drive family "A" from supplier "X" may have a specific problem that most often shows up as "04-wx-yz". However, drive family "B" from supplier "Y" may have a different problem that results in the same error code, "04-wx-yz". Due to this vendor specific mapping between error codes and mechanisms, the modes can often be treated as synonymous with the mechanism for specific drive families and specific error codes. A list of dominant failure mechanisms that occur in disk drives today is presented as Table 3.

Corrosion	Media contamination
Thermal Erasure	Particle under head
Media Errors	T/A: Head hits bump Defect in media (embedded in mfgr.) High-fly writes Rotational vibration Hard particles ("loose") Head slaps (shock & vib.)
Head Instability (& Dead Head)	Repeated T/A's Head assymetry in mfgr, Many, long writes Hard particle contact ESD
Hardware Errors	Motors, bearings; design, mfgr. PWA: solder, connectors
Electrical Failures	PWA: DRAM, ICs PWA: Motor drivers, pre-amp. chip
Acronyms:	
T/A	Thermal asperity
ESD	Electrostatic Discharge
PWA	Printed Wiring Assembly
DRAM	Dynamic Random Access Memory
IC	Integrated Circuit

Table 3. Disk Drive failure Mechanisms

However, for proprietary reasons, direct mappings between mechanisms and error codes will not be presented in this paper. The charts presented as figures 2 through 9 are, in truth, based on error codes, but because of the unique mapping, are synonymous with mechanisms from each specific supplier. The failure mechanisms (error codes) are identified as FM1, FM2 etc, and the X and Y-axis values on the charts hidden to preserve the proprietary and sensitive nature of the data.

4. *ANALYSES*

Disk drive manufacturers often create what are called drive "families". Within a drive family, the drives are nearly identical except for changes such as the number of disks and read/write heads. Families are used to maximize product commonality, which reduces design and manufacturing costs and addresses the need of the market for multiple capacity points, such as 36GB, 72GB and so on. Each family of drives goes through different design and manufacturing processes. Some failure mechanisms are independent of the capacity of the drives within the drive family, while others are dependent on the number of read/write heads and the number of disks. Reliability analyses are performed on three such disk drive families from different drive manufacturers to study the effect of failure mechanisms on the overall hazard rate of disk drives.

4.1 *Effect of Different Failure Mechanisms on Overall Hazard Rate*

Figure 3 is the hazard rate plot of different failure mechanisms of drive family A. Hazard rate (failure rate as a function of time) is calculated using formula:

$$h(t)=\frac{f(t)}{R(t)} \tag{1}$$

Where, $f(t)$ is the probability density function and $R(t)$ is reliability as a function of time. Hazard rate is plotted on the ordinate, whereas the time is plotted on the abscissa. The composite hazard rate of all failure mechanisms combined appears to be increasing in the beginning, followed by a constant failure rate, which is then followed by a slightly increasing hazard rate. Drive A (Figure 3) seems to have at least three different failure mechanisms at work, infant mortality, constant failure rate and an increasing failure rate.

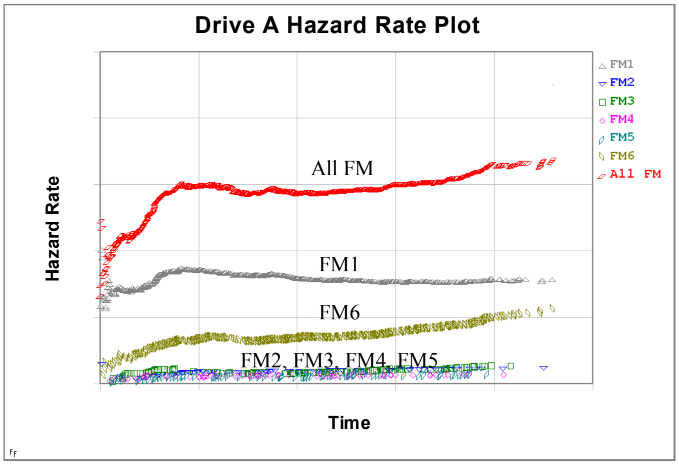

Figure 3. Hazard Rate Plot of Failure Mechanisms of Drive A

To understand the cause of this changing failure rate nature, a hazard rate plot of different failure mechanisms contributing to the overall hazard rate is plotted individually.

The hazard rate of failure mechanism FM1 is slightly increasing in the beginning and then decreasing. This failure mechanism is related to the infant mortality failures caused by handling or particulate contamination, which dominate in the early period of the drives operating life, but decreases as drive manufacturing processes mature. Failure mechanisms FM2, FM3, FM4 and FM5 are rather constant throughout the operating life of the drive. These represent the random failures, which dominate during the useful life of the drive. Failure mechanism FM6 seems to have a gradually increasing hazard rate and starts to dominate the overall hazard rate towards the later part of the drives operating life. Disk drives with failure mechanism FM6 are identified and a corrective action plan is implemented to mitigate the overall hazard rate of drives in the field.

4.2 *Effect of Firmware Changes Dependent Failure Mechanisms on Overall Hazard Rate*

Disk drive manufacturers often introduce new drive firmware versions to fix certain "firmware fixable" failures in the field during the production cycle of the disk drive. These changes generally do not involve major design effort and are relatively easier to incorporate during the production of the drive than the hardware design changes. This analysis studies the effect of the failure mechanisms that are dependent on these firmware changes.

Figure 4 is the hazard rate plot of different failure mechanisms of drive family B. The composite hazard rate of all failure mechanisms combined is increasing from the beginning. The hazard rate of failure mechanisms FM1 and FM5 are almost constant in the beginning part of drives operating life and increases rapidly after a certain period of time in the field.

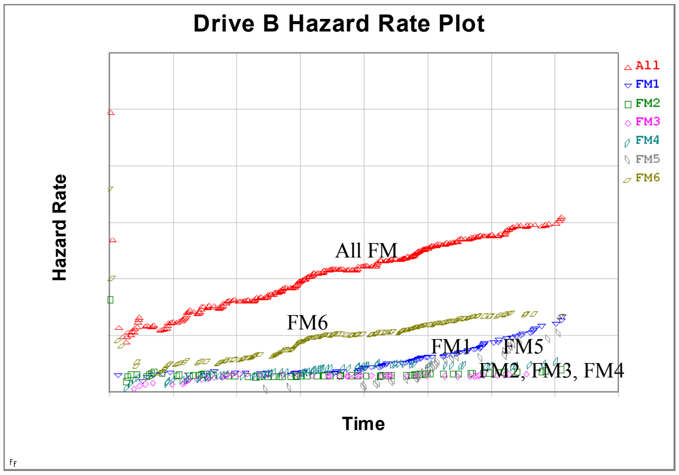

Figure 4. Hazard Rate Plot of Failure Mechanisms of Drive B

After much investigation into the root cause of failure mechanisms FM1, FM5 and FM6, it was found that a drive firmware "X" was introduced into the field to fix a particular problem around the same time the hazard rate of failure mechanisms FM1, FM5 and FM6 started to increase. The changes that were made in the drive firmware X were actually accelerating these failure mechanisms at a very high rate causing the increase in the hazard rate.

To better understand the effect of drive firmware version X, the population of drive family B is segregated into two sub-populations. First population representing the drives operating on the firmware version <X and the second population representing the drives operating on firmware version >=X. These two subpopulations hazard rates are then plotted as shown in figures 5 and 6.

Figure 5 depicts the hazard rate of all failure mechanisms before the firmware version X was introduced. The hazard rate of all failure mechanisms is very low and almost constant.

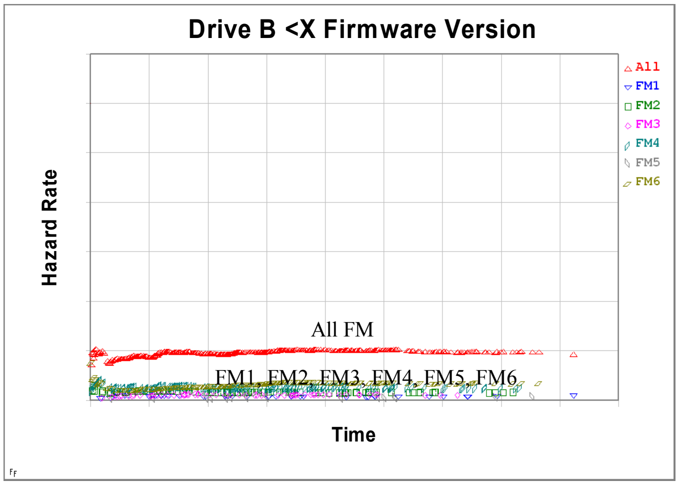

Figure 5. Hazard Rate Plot of Failure Mechanisms of Drive B Before Firmware Version X

Figure 6 depicts the hazard rate of all failure mechanisms of drives operating after the firmware version X was introduced.

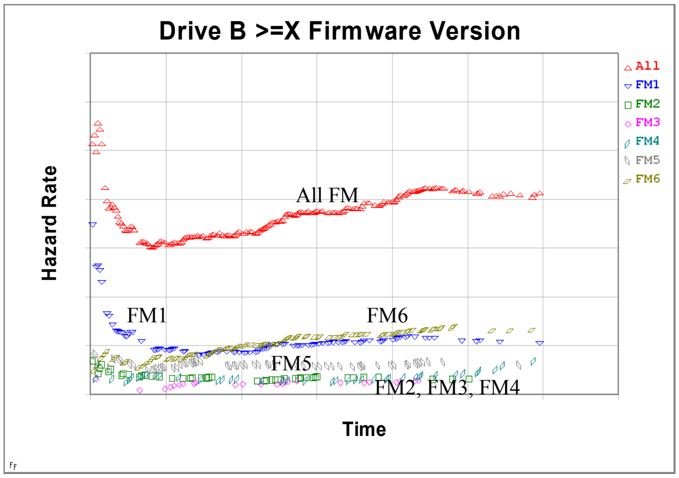

Figure 6. Hazard Rate Plot of Failure Mechanisms of Drive B After Firmware Version X

The hazard rate of failure mechanisms FM1, FM5 and FM6 are much higher and gradually increasing. The composite hazard rate due to the combination of the hazard rates of individual failure mechanisms of drives operating on firmware version >=X was found to be almost 4 times higher than the composite hazard rate of drives operating on firmware version before X was introduced. This type of analysis could further be extended to understand the effect of the changes made in the operating system software.

4.3 *Effect of Capacity Dependant Failure Mechanisms on Overall Hazard Rate*

Most disk drive families have different capacities. These drives go through the same design and manufacturing process. The main difference is the number of read/write heads and disks. This analysis studies the effect of the failure mechanisms that are dependent on capacity of the drive. Figure 7 is a hazard rate plot of two capacities of drive family C and the overall hazard rate of all drive capacities combined plotted on a linear plot. The high capacity drives seem to have an increasing hazard rate from the beginning, whereas the low capacity drives seem to have a rather constant hazard rate. Therefore, the composite hazard rate of both low capacity and high capacity drives is gradually increasing. To understand the difference in the hazard rates between the high and low capacity drives, hazard rate plots of all failure mechanisms are plotted on separate plots.

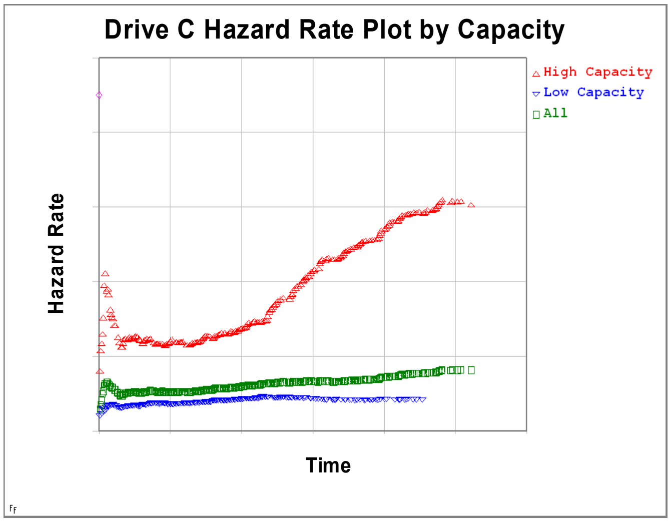

Figure 7. Hazard Rate Plot of Different Capacity of Drive C

Figure 8 is a hazard rate plot of low capacity drives of drive family C. All failure mechanisms seem to have a rather constant and low hazard rate so the composite hazard rate is also constant.

Figure 9 is the hazard rate plot of different failure mechanisms of high capacity drives. All failure mechanisms except failure mechanism FM2 have a higher hazard rate compared to low capacity drives. Additionally, failure mechanisms FM4 and FM5 seem to have an increasing hazard rate. After much root cause and failure analysis on these drives, it was found that the majority of the failures mechanisms were related to the particulate contamination induced during the manufacturing process. This failure mechanism was highly prominent for the high capacity drives because the high capacity drives have more read/write heads and more disks per drive than the low capacity drives and consequently more opportunities of defects per drive.

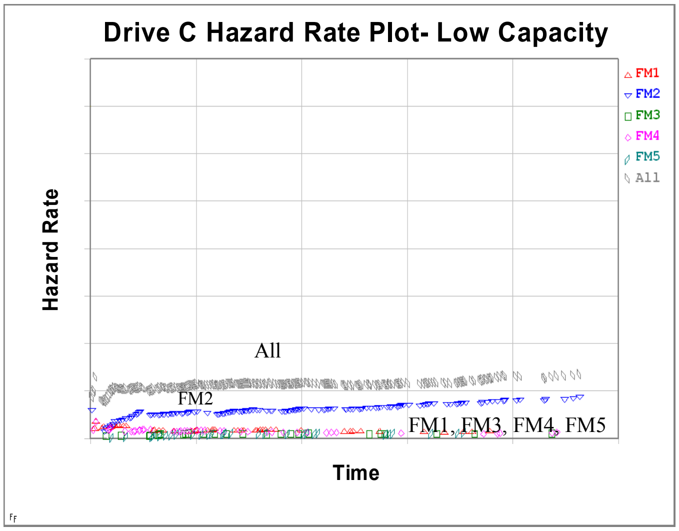

Figure 8. Hazard Rate Plot of Failure Mechanisms of Low Capacity of Drive C

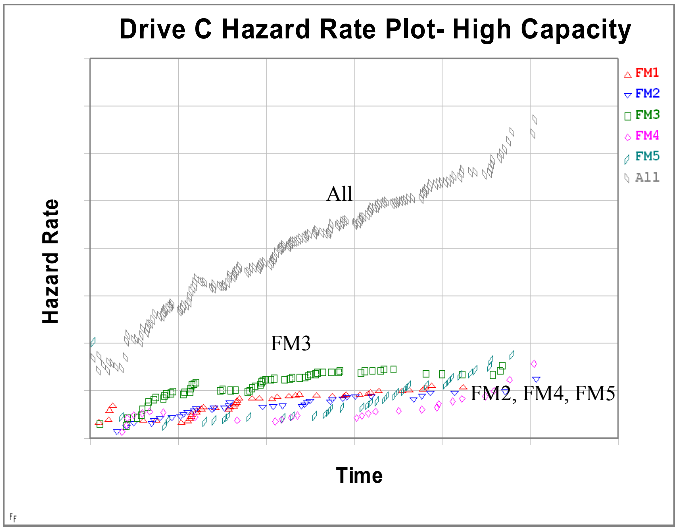

Figure 9. Hazard Rate Plot of Failure Mechanisms of High Capacity of Drives C

The results of these analyses illustrate that disk drive reliability is highly dependent on the manufacturer and family of disk drive. The most prevalent cause of failure in general is a result of contamination. Some contaminants are buried in the media coatings on top of the disk substrate resulting in thermal asperities, while others are “loose” particles that cause scratches and damaged heads. Head stability is also a critical reliability detractor. Drive level reliability measurements are unable to identify these causes, but good failure analysis and a good reliability database allow time dependent analyses on the failure mechanisms. These analyses are used to work with the supplier to quantify the magnitude of the problem.

REFERENCES

1. IDEMA Standards, *Specification of Hard Disk Drive Reliability*, document number R2-98.
2. Abernathy, Robert B. *The New Weibull Handbook*, Fourth Edition, November 2000.
3. Elerath, J. G., and Shah, S., "Disk Drive Reliability Case Study: Dependence Upon Head Fly-Height and Quantity of Heads," *Proc. Annual Reliability & Maintainability Symp.,* January 2003.
4. Shah, S., and Elerath, J. G., "Disk Drive Vintage and Its Effects on Reliability," *Proc. Annual Reliability & Maintainability Symp.,* January 2004.

BIOGRAPHIES

Sandeep Shah
Network Appliance
495 East Java Drive
Sunnyvale, CA 94089 USA

e-mail: sandeep.shah@netapp.com

Sandeep Shah has a BS in Mechanical Engineering from Osmania University, India and an MS in Reliability Engineering from the University of Arizona. He is currently working at Network Appliance Inc., as Reliability Engineer. He has co-authored papers on disk drive reliability, structural reliability and Finite Element Analysis. He has 7 years of experience in the field of Reliability in robotics (PRI Automation), power tools (Black & Decker) and data storage systems (Network Appliance).

Jon G. Elerath
Network Appliance, Inc.
495 E. Java Drive
Sunnyvale, CA 94089 USA

e-mail: jon.elerath @netapp.com

Jon Elerath has a BSME and MS in Reliability Engineering from the University of Arizona. He has 29 years experience in reliability engineering and engineering management in areas of nuclear safety systems (General Electric Co.), plasma etching equipment for semiconductor manufacturing (Tegal, a subsidiary of Motorola), fault tolerant computers (Tandem and Compaq), hard disk (Winchester) drives (IBM) and storage systems (Network Appliance). He is currently the manager of Reliability Engineering for Network Appliance Inc., a world leader in network-attached storage systems. His has participated in all aspects of reliability in a commercial environment, including developing overall reliability programs, specifications, predictions, trade-off analyses, testing and data collection and analysis. He has chaired the Redwood Empire Section of ASQ, the Reliability Committee of the International Disk Drive Equipment Materials Association (IDEMA) and has contributed greatly to the development of the IEEE 1413 standard for reliability predictions. He has published 20 papers in the area of reliability and reliability modeling.