



THIC Inc.

The Premier Advanced Recording Technology Forum

MAID for Archiving

Aloke Guha

COPAN Systems

1900 Pike Road, Longmont, CO 80501-6775, US

Phone: +01 303-532-0250 FAX: +01 303-532-0302

E-mail: aloke.guha@copansys.com

**Presented at the THIC Meeting at the National Center for Atmospheric
Research, 1850 Table Mesa Drive, Boulder CO 80305-5602**

July 19-20, 2005

Archive

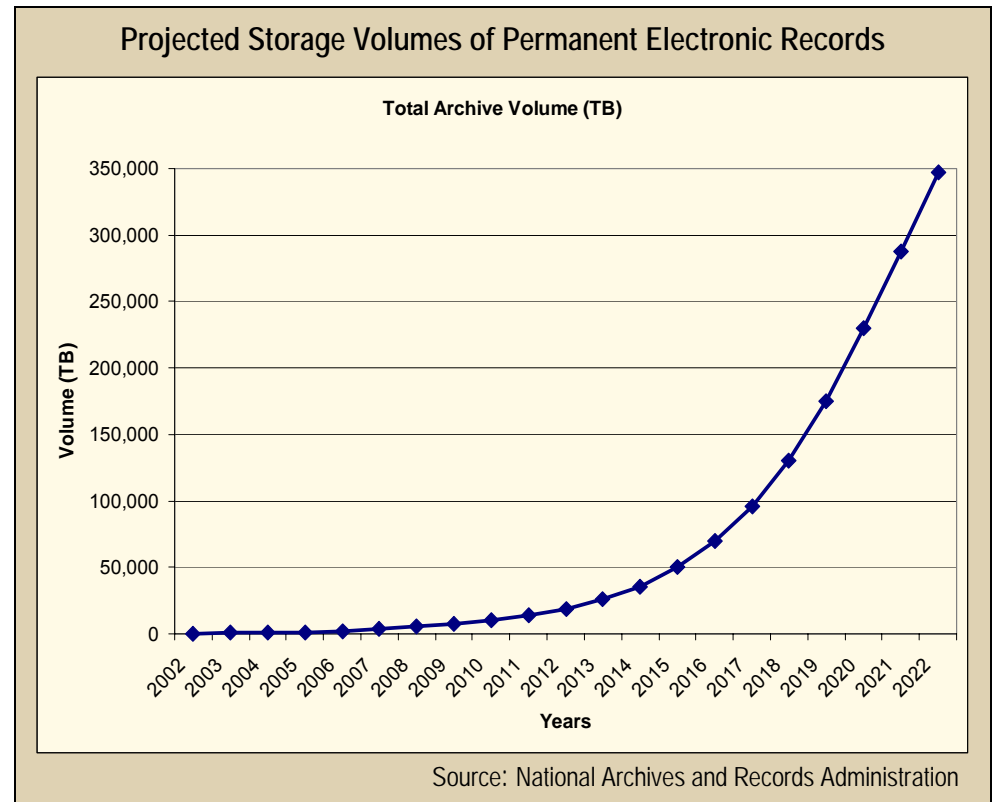
- **Definition:** *An archive is a collection of data that is maintained as a long-term record of a business, an application, or an information state[#]*
- **Uses**
 - **Auditing, Regulatory, Analysis, Fixed Content, Reference Data**
- **Types**
 - **Active Archives:** frequent access to data, up to many times/day, many accesses to data within same data object
 - **Deep Archive:** long-term vaulted data that have very infrequent access, more tolerant of large latencies

[#]ILM Best Practices in Archiving and Tiered Storage for Email, SNIA Data Management Forum, Oct. 2004

Archiving: More than Storage

- **Archive Storage**
 - **Scale – growing capacity**
 - **Cost effective**

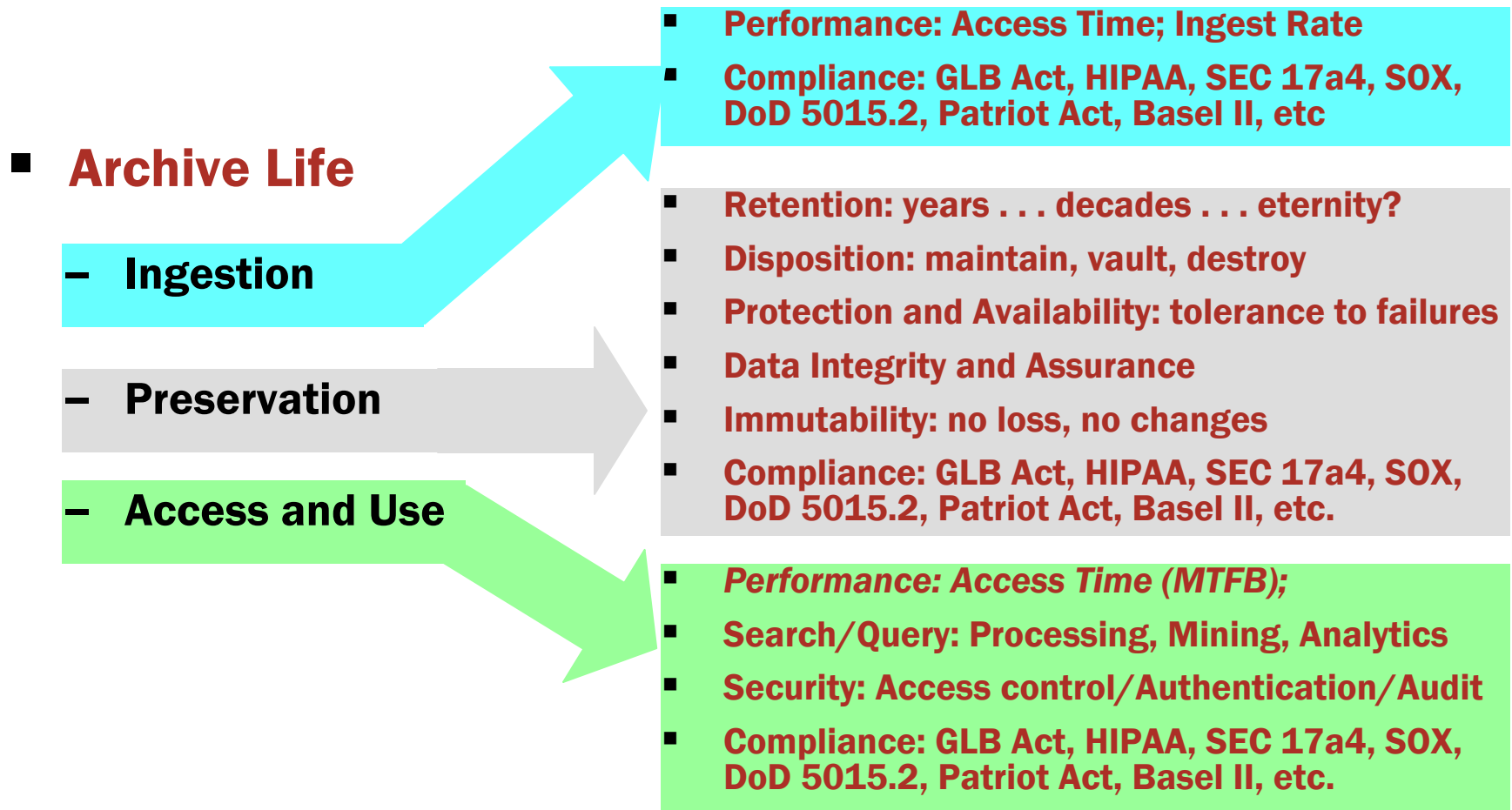
- **Archive Life . . . beyond storing bits**
 - **Ingestion – how data is entered**
 - **Preservation – maintenance and health of records[#]**
 - **Access and Use – how records are retrieved**



[#] K. Thibodeau, Building the Archives of the Future, Gartner PlanetStorage June 2005

[#] M. Smith, Eternal Bits, IEEE Spectrum, July 2005

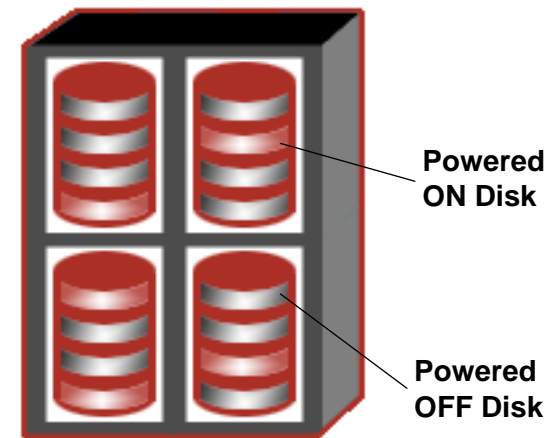
Archiving: Attributes . . . More than Storage



. . . and Cost-Effective and Scalable Storage

MAID: Power-Managed Storage System

- **Large # power-managed disks**
 - > 50% drives powered OFF
 - Power-cycling by policy
- **Scale, Cost, Service Life**
- **Lower System Cost/Drive**
 - 1/4 - 1/3 of typical RAID systems
 - Lower management cost
- **Extending MAID**
 - Performance and Scale
 - Reliability
 - Cost
- **Applicability**
 - Backup and Recovery
 - Long- Term Data: Archive



SNIA Definition#

“A storage system comprising an array of disk drives that are powered down individually or in groups when not required. MAID storage systems reduce the power consumed by a storage array.”

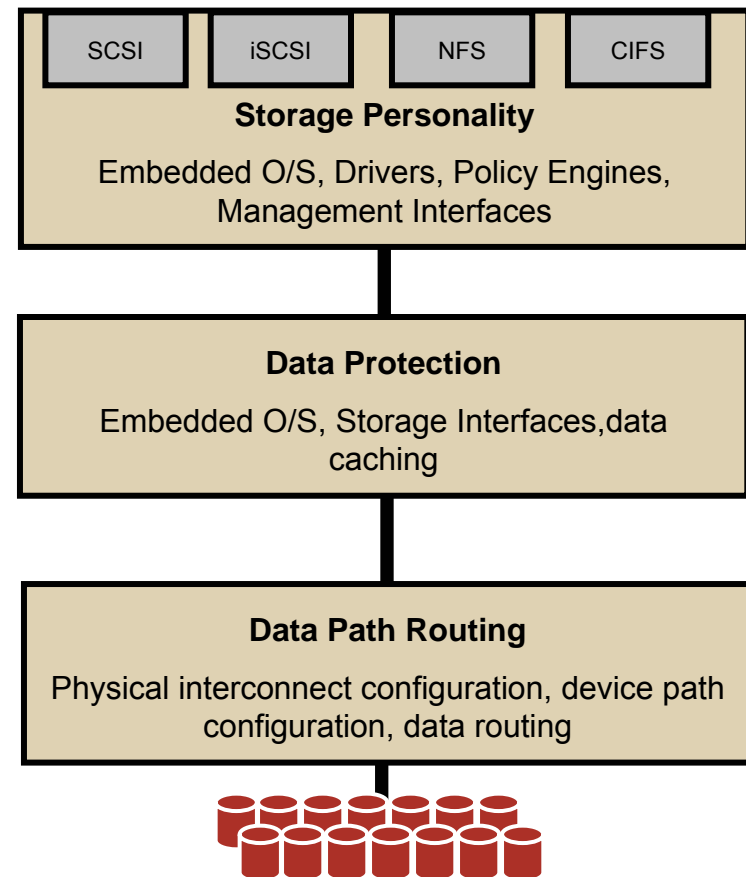
COPAN Architecture: 3-Tier MAID Architecture

Three levels of processing separates functionality, simplifies management, scales performance w/ capacity

- **Layer 2 – Storage Personality**
 - **Physical Domain (Revolution Controller)**
 - **Storage Network Protocols**
 - **Logical Volume/Block Management**
 - **Performance and Load Balancing**

- **Layer 1 – Data Protection**
 - **Physical Domain (Shelf Controller)**
 - **RAID Support and Caching**
 - **Power Management**
 - **Device Management**

- **Layer 0 – Data Path Routing**
 - **Physical Domain (Canister Controller)**
 - **Protocol Router**
 - **Monitoring of environmental parameter**



Architecture: Tuned for Long-Term Data

Accessible, Retained and Protected

- **Architecture: 3-Tier MAID Architecture[#]**

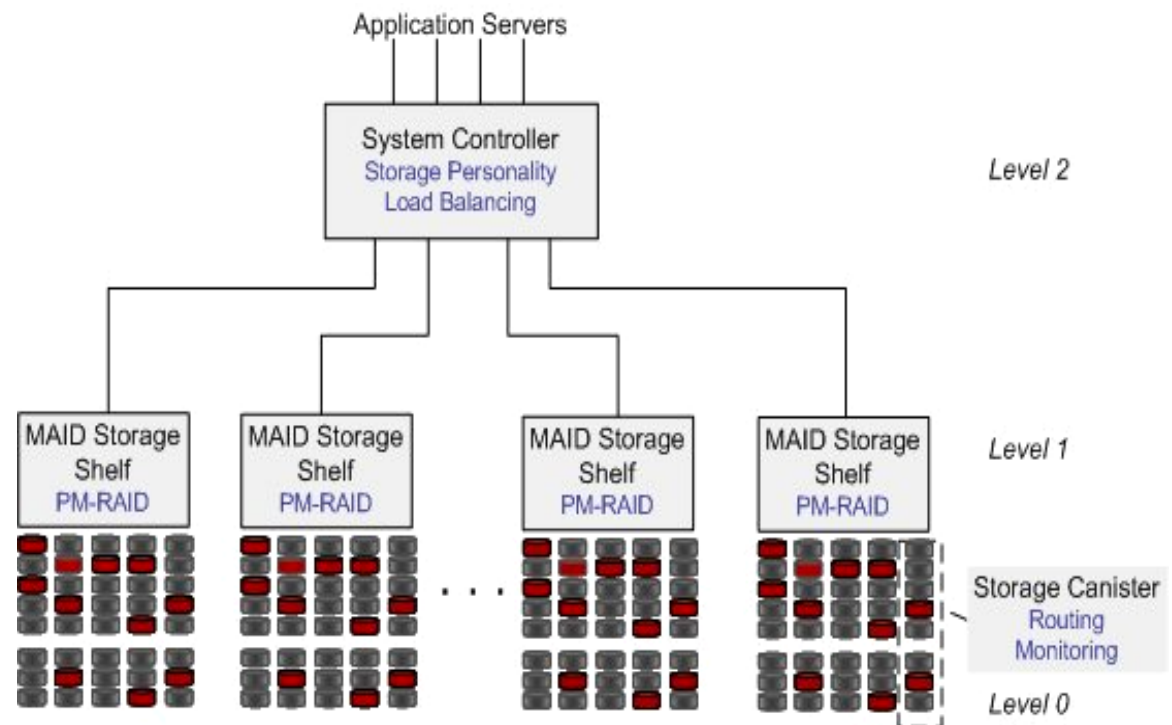
- Storage Density
- Scale
- Performance
- Serviceability
- Power/Cooling

- **Data Reliability**

- **POWER MANAGED RAID™ Software**

- **Data Retention**

- **DISK AEROBICS™ Software**



[#]A. Guha, A New Approach to Disk-Based Scalable Mass Storage System, IEEE Conf. Mass Storage Systems, April 2004

COPAN's Revolution 200T

8 shelves/system
8 canisters/shelf
14 drives/canister

(896) 250 GB drives
= **224 TB**
per system!
(uncompressed)

30"W x 48"D x 84"H
(44U)

= 22TB/sq.ft.

4 –FC Connections

Performance

= 2.75TB / HR



Three-Tier System Architecture

- Three levels of processing separate functionality
- Scale performance with capacity
- Simple management

POWER MANAGED RAID™ Software

- RAID 5 data protection
- Drives spin only when necessary to meet application requirements
- <= 25% drives powered: extends service life

DISK AEROBICS™ Software

- Predictive monitoring and management of drive health
- Periodically exercises and monitors non-powered on drives
- Proactively detecting possible drive failures before they occur

MAID and Archiving

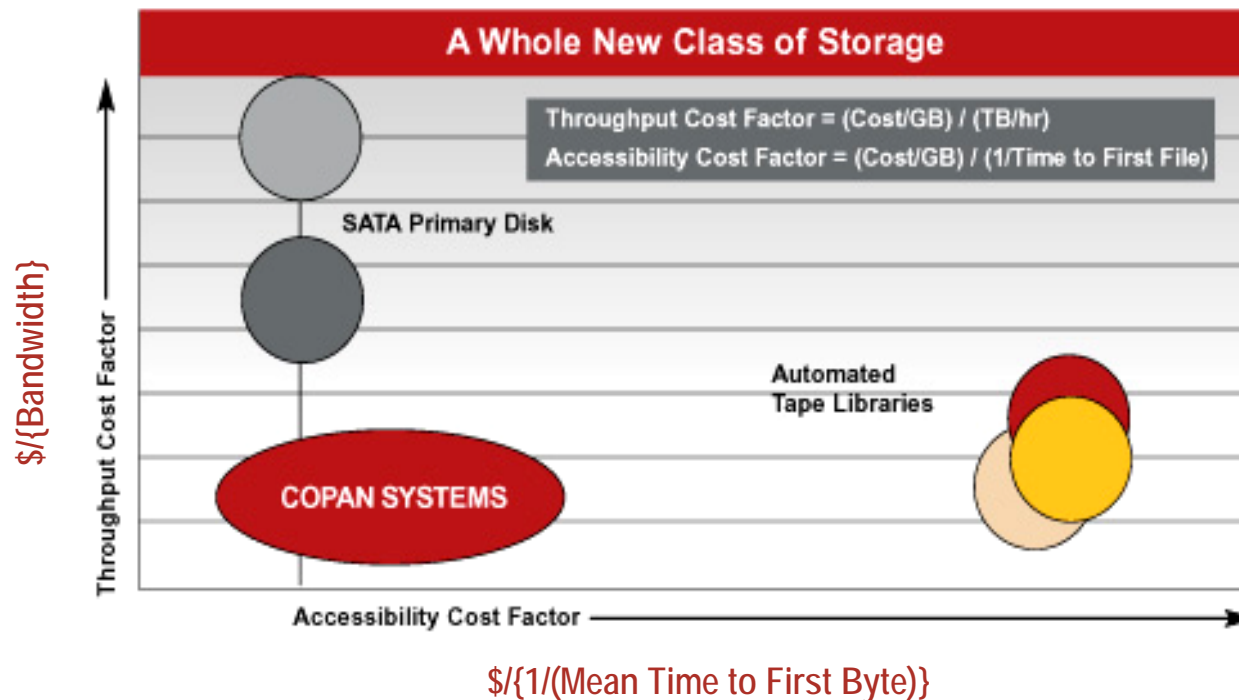
- **Scale and Capacity: Storage Density and Scale**
 - MAID Scale Optimized
 - Cost Optimized
- **Ingest: Data Transfer Rate, Latency**
 - Performance Tuned to Workload - Why “Burn” 100% drives for <25% Load
 - Process (Compliance): Logical Presentation - 3-Tier Architecture Control
- **Preservation**
 - Retention (Compliance): DISK AEROBICS™
 - Data and Storage Management
 - Data Integrity: DISK AEROBICS™ -- verify data within last 30 days
 - Cost-Effective Protected Data: POWER MANAGED RAID™
 - Immutability (Compliance): Read-Only – 3-Tier Architecture Control
- **Access and Use**
 - Performance: MTFB, Data Transfer Rate
 - Security: Access Control/Authent./Authorization – 3-Tier Architecture

Storage Density and Scale

- **Storage density (drives/vol) driven by power duty-cycle (D)**
 - Density increases with decreasing temperature
 - Density increases with decreasing vibration
- **Interconnect options: # of drives vs. # of ports**
 - Cost of switching increases with # of ports per single switch
 - Multiple stages meets with cost and practical technology
- **Result**
 - 3-Tier interconnect: {#shelves s, #canisters c, #drive d/canister}
 - Optimization: {dim (s, c, d), Power/Cooling, Serviceability, Redundancy, Bandwidth, Processing/App Support} wrt Cost
 - 896 drives/cabinet: max 25% of user data online ($D \leq 0.27$)
 - 250 GB drives: 224 TB Uncompressed; 22.4 TB/sq ft
 - 500 GB drives: 448 TB Uncompressed; 44.8 TB/sq ft

Ingest: Performance

- **Bandwidth: scales with multiple RAID engines**
 - Write rate/frame: 2.75 TB/hr (limited by # ports)
- **Access Time**
 - Average first time to access data on spin up: 0 (secs)
 - Subsequent (non-sequential) access to data: ~ 0 (ms)

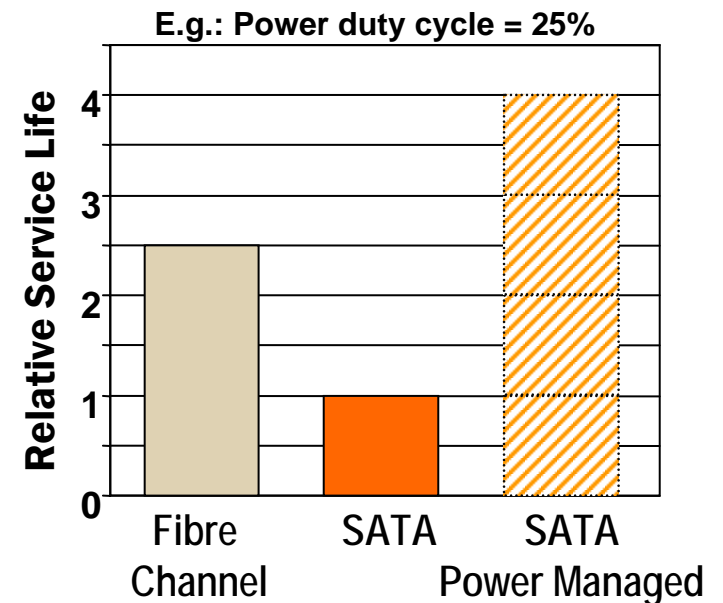


Access and Use: Performance Close to Disk

- **Retrieve Rate**
 - Read rate/frame: 2.75 TB/hr (limited by # ports)
- **File Transaction Throughput**
 - Performance approaches that of standard disk
 - If locality/volume increases 100X => throughput increases 100X
- **Provides option of archive data processing not possible with tape**
 - Performance orders of magnitude higher than tape or tape+staging disk
 - Effectively on-line data for data mining, queries, etc.
- ***Most cost-effective solution for active archives!***

COPAN MAID and Data Reliability

- **Effective drive service life**
 - Increases with decreasing duty ratio*
- **Proactive monitoring: DISK AEROBICS™**
 - Proactive data recovery before drive failure
 - Rebuild data transparently
- **Increases data reliability**
- **Explicitly manage start stops**
 - ≤ 50K over service life
- **Exploit disk density for availability**
 - Spares to replenish failed drives
 - *Data Revitalization for long-term data*



Revitalized Data + Drive Monitoring ⇒ Longer Data Retention

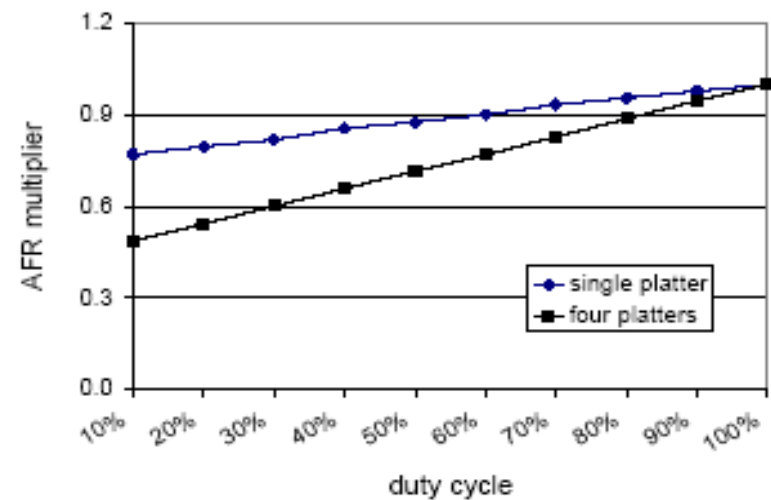
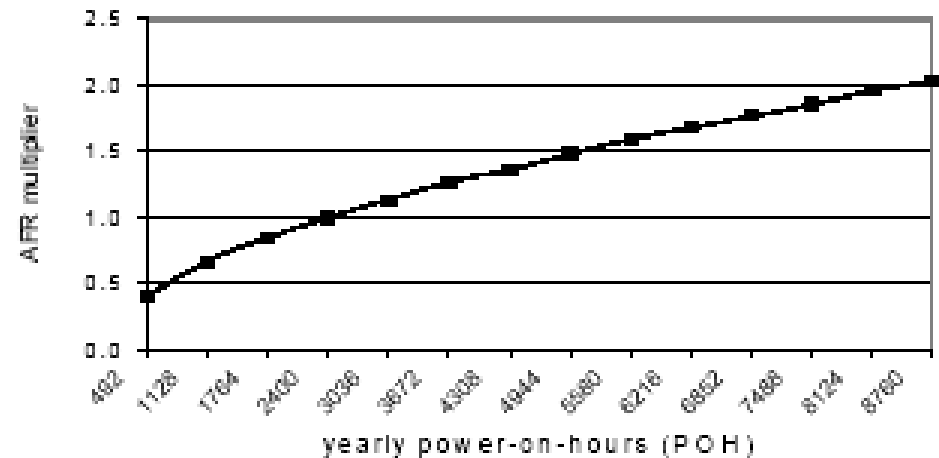
Disk and Data Reliability with MAID

- **Disk Reliability and Duty Cycle**
 - Decreasing POH: decreases annual failure rate (AFR), increases reliability
 - Decreasing duty cycle: decreases AFR, increases reliability

- **Archiving and Drive Reliability**
 - “Power-off archival storage stresses fewer spin motor bearing failure issues”
 - “In disk drive archives, each drive should be periodically powered up in a MAID system, or mounted in a drive tester”

Hughes, Murray, “Reliability and Security of RAID Storage and D2D Archives Using SATA Disk Drives, ACM Trans. of Storage, Dec 2004

Anderson, et al, “More than an Interface: SCSI vs ATA”, 2nd Annual Conf. on File and Storage Tec, Mar 2003



Conclusions

- **Archive needs go well beyond storing bits**
 - Ingest – Preserve – Access/Use
- **Tuned MAID Optimized for Archives**
 - Capacity Scale, Storage Density: MAID, 3-Tier Architecture
 - Ingest: Online Data Performance
 - Preservation: POWER MANAGED RAID™, DISK AEROBICS™, 3-Tier Architecture (Immutability)
 - Access/Use: Online Data Performance, 3-Tier Architecture (Data Processing)
- **Most Cost-Effective Solution of Active Archives!**



Thank you

<http://www.copansys.com>