

NASA Conference Publication 3165, Vol. I

NSSDC Conference on Mass Storage Systems and Technologies for Space and Earth Science Applications

Volume I

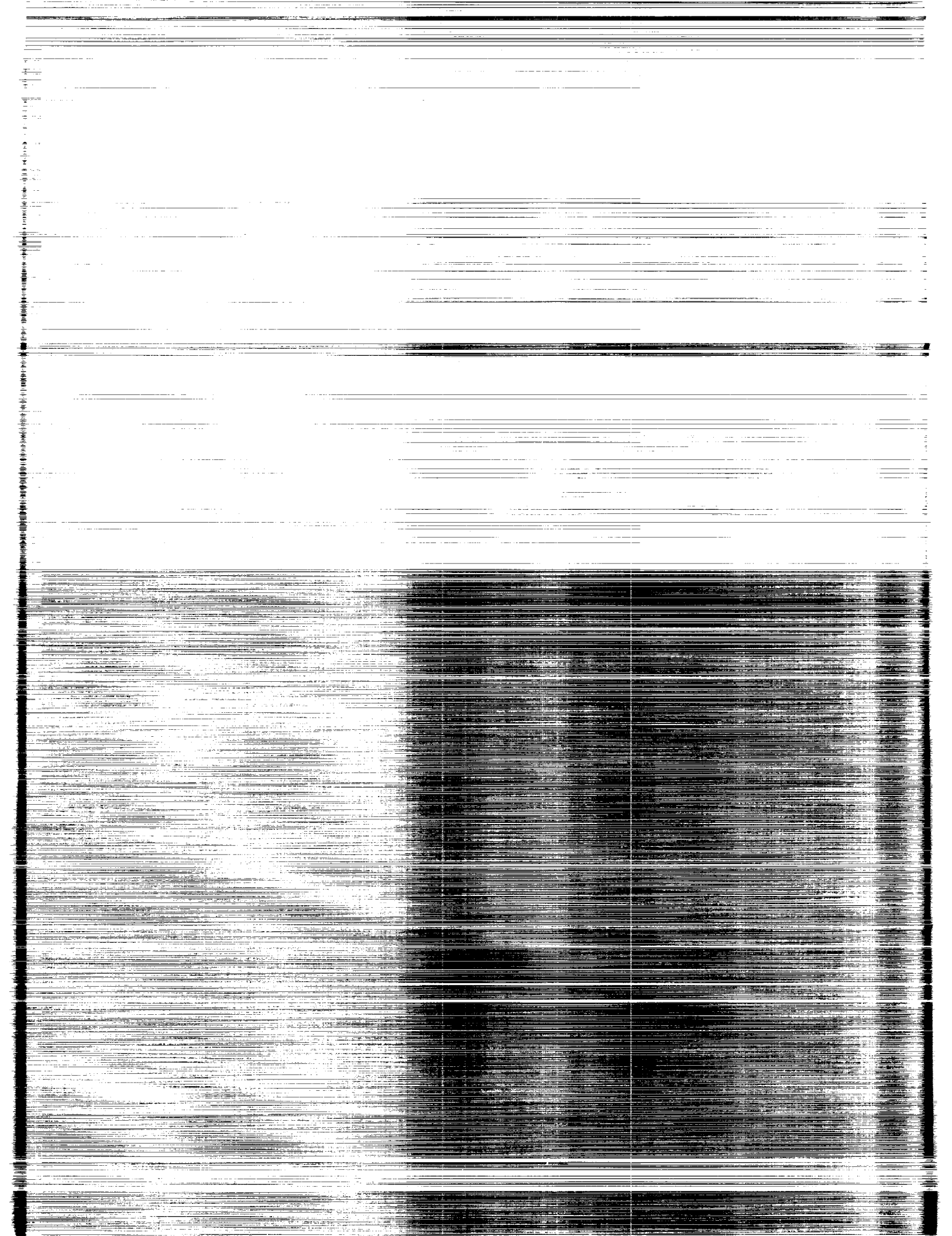
*Proceedings of a conference held at
NASA Goddard Space Flight Center
Greenbelt, Maryland
July 23-25, 1991*

(NASA-CP-3165-Vol-1) NSSDC
CONFERENCE ON MASS STORAGE SYSTEMS
AND TECHNOLOGIES FOR SPACE AND
EARTH SCIENCE APPLICATIONS, VOLUME
1 (NASA) 205 p

N93-13131
--THRU--
N93-13139
Unclas

H1/82 0122370

NASA



NASA Conference Publication 3165, Vol. I

NSSDC Conference on Mass Storage Systems and Technologies for Space and Earth Science Applications

Volume I

*Edited by
Ben Kobler
Goddard Space Flight Center
Greenbelt, Maryland*

*P. C. Hariharan
and L. G. Blasso
STX Corporation
Lanham, Maryland*

Proceedings of a conference held at
NASA Goddard Space Flight Center
Greenbelt, Maryland
July 23-25, 1991

NASA

National Aeronautics and
Space Administration
Office of Management
Scientific and Technical
Information Program

1992

Preface

The National Space Science Data Center (NSSDC) at NASA's Goddard Space Flight Center has been charged with the archiving of data collected from NASA's scientific spaceflight missions flown over the past 30 years. During this time NSSDC has accumulated an archive of several terabytes of data. In the coming years NASA will be generating this volume of data every few days or less. Thus, data storage media and systems become critically important to NASA if it is to successfully manage this data volume and to have a chance to transform these data into scientific knowledge.

NSSDC will play an important role in NASA's awareness of and exploitation of emerging mass storage systems, both at NSSDC and in the increasingly distributed NASA scientific data environment. For this reason, NSSDC organized a conference at Goddard in the summer of 1991 to review the status of and the outlook for data storage media and systems. Leading experts in each of several areas were invited to make presentations, and a highly informative conference transpired. In order that the record of that conference be preserved, this set of presentations is being published.

The Proceedings of the NSSDC Conference on Mass Storage Systems and Technologies for Space and Earth Science applications are published in four volumes, with each of the first three volumes containing the talks and presentations for that particular day. Discussions following some of the talks are collected in the fourth volume along with introductory biographical material on the speakers. Despite our best efforts, the questions and answers were sometimes inaudible to the transcriptionist. An effort was made to contact the participants to clarify the transcript, and we are grateful to the speakers who cooperated.

The success of an endeavor of this magnitude depends on the generous help and cooperation of the participants. We would like to record our gratitude to the speakers, the audience, and in particular, to the following individuals and organizations for their assistance:

The Program committee whose membership is listed in the front of each volume

The session chairs who kept the schedule on track:

Professor Bharat Bhushan of Ohio State University
Dr. Barbara Reagor of Bellcore
Dr. Robert Freese of Alphatronix
Mr. Patric Savage of Shell
Professor John C. Mallinson of Mallinson Magnetics
Dr. Kenneth Thibodeau of the National Archives and Records Administration

The members of the Panel Discussion and Professor Mark Kryder of Carnegie Mellon University who moderated the discussion

The hard-working crew from Westover Consulting

Dr. Dennis E. Speliotis of Advanced Development Corporation for his help with the transcript of the Panel Discussion

Dr. James L. Green and the National Space Science Data Center

Dr. J. H. King
Dr. P. C. Hariharan
Benjamin Kobler

NSSDC Mass Storage Conference Committee

Ben Kobler, *NASA/GSFC (Chair)*
John Berbert, *NASA/GSFC*
Sue Kelly, *DOE/Sandia*
Elizabeth Williams, *SRC/Bowie*
Al Dwyer, *STX Corporation **
P. C. Hariharan, *STX Corporation **
Sanjay Ranade, *STX Corporation **

Conference Coordinator

Kim Blackwell, *STX Corporation **

* Hughes STX Corporation as of October 1, 1991

MASS STORAGE CONFERENCE PROCEEDINGS

CONTENTS

VOLUME I - FIRST DAY - TUESDAY, JULY 23, 1991

1. *Introduction*..... 1-1
Dr. Milton Halem, Chief
NASA Space Data and Computing Division
2. *Enterprise Storage Report for the 1990s*..... 1-3
Fred Moore, Corporate Vice President, Strategic Planning
Storage Technology Corporation
3. *Optical Disk and Tape Technology*..... 1-33
Dr. Robert Freese, President, Alphatronix
4. *Magnetic Disk*..... 1-69
Dr. John C. Mallinson, Mallinson Magnetics, Inc.
5. *Magnetic Tape*..... 1-79
L. Harriss Robinson, Manager, Mass Storage Systems
Datatape, Inc.
6. *Storage System Software Solutions For High-End User Needs*..... 1-109
Carole Hogan, Director of Technical Development, DISCOS
7. *File Servers, Networking and Supercomputers*..... 1-145
Dr. Reagan Moore, San Diego Supercomputing Center
8. *Mass Storage System Experiences and Future Needs at National Center
for Atmospheric Research*..... 1-163
Bernard T. O'Lear, Manager, Systems Programming,
Scientific Computing Division, NCAR
9. *The Long Hold: Storing Data At The National Archives*..... 1-183
Dr. Kenneth Thibodeau, Director, Center for Electronic Records,
National Archives and Records Administration
10. *Banquet Presentation*..... 1-187
Dr. John C. Mallinson, former Director of Center for Magnetic Recording
Research, University of California, San Diego

MASS STORAGE CONFERENCE PROCEEDINGS

CONTENTS (Continued)

VOLUME II - SECOND DAY - WEDNESDAY, JULY 24, 1991

1. *Stewardship of Very Large Digital Data Archives*.....2-1
Patric Savage, Shell Development Company
2. *High-Performance Storage Systems (Text Not Included in this Publication)*.....2-7
Robert Coyne, IBM Federal Systems Research Division
3. *An Open, Parallel I/O Computer as the Platform for High-Performance,
High-Capacity Mass Storage Systems*.....2-9
Adrian Abineri, APTEC Computer Systems
4. *EMASSTM: An Expandable Solution for NASA Space Data Storage Needs*.....2-21
Anthony L. Peterson, P. Larry Cardwell, E-Systems, Inc.
5. *Data Storage and Retrieval System*.....2-35
Glen Nakamoto, MITRE Corporation
6. *The Challenge of a Data Storage Hierarchy: Data Archiving*.....2-55
Michael Ruderman, Mesa Archival Systems, Inc.
7. *Network Accessible Multi-Terrabyte Archive*.....2-85
Fred Rybczynski, Metrum Information Storage
8. *ICI Optical Data Storage Tape*.....2-98
Robert A. McLean, Joseph F. Duffy
9. *ATL Products Division's Entries into the Computer Mass Storage Marketplace*.....2-110
Fred Zeller, Odetics Automated Tape Library (ATL)
10. *Panel Discussion*.....2-122
Moderator: Dr. Mark Kryder
Dr. Barbara Reagor, Bellcore
Darlene M. Carlson, 3M National Media Laboratory
Kazuhiro Okamoto, Sony Magnetic Products, Inc.
Allan S. Hadad, Ampex Recording Media Corporation
John W. Corcoran, Consultant/Ampex Corporation
Dr. Dennis E. Spiliotis, Advanced Development Corporation
11. *Poster Session*.....2-237
Steve Atkinson, Ampex Recording Media Corporation
Barbara Matheson, STX Corporation
Kirby Collins, CONVEX Storage Systems

MASS STORAGE CONFERENCE PROCEEDINGS

CONTENTS (Continued)

VOLUME III - THIRD DAY - THURSDAY, JULY 25, 1991

1. *Tribology of Magnetic Storage Systems*..... 3-1
Dr. Bharat Bhushan, Ohio State University
2. *Network Issues for Large Mass Storage Requirements*3-73
James "Newt" Perdue, Ultra Network Technologies, Inc.
3. *The Role of HIPPI Switches in Mass Storage Systems; A Five Year Prospective*...3-97
T. A. Gilbert, Network Systems Corporation
4. *The National Space Science Data Center: An Operational Perspective*.....3-127
Ronald Blitstein, Hughes STX Corporation/NSSDC
Dr. James L. Green, NSSDC
5. *EOSDIS/DADS Requirements*.....3-141
John Berbert, Ben Kobler
6. *The Preservation of Landsat Data by the National Land Remote Sensing
Archive*.....3-153
John E. Boyd, EROS Data Center
7. *Status of Emerging Standards for Removable Computer Storage Media and
Related Contributions of NIST*.....3-163
Fernando L. Podio, National Institute of Standards and Technology Computer
Systems Laboratory
8. *Data Management in the National Oceanic and Atmospheric Administration
(NOAA)*.....3-189
William M. Callicott
9. *Storage Needs in Future Supercomputer Environments*3-211
Dr. Sam Coleman, Lawrence Livermore National Laboratory
10. *Requirements for a Network Storage Service*3-225
Suzanne M. Kelly, Rena A. Haynes, Sandia National Laboratories

MASS STORAGE CONFERENCE PROCEEDINGS

CONTENTS (Continued)

VOLUME III - TRANSCRIPTIONS

<i>Introduction of Fred Moore.....</i>	<i>3-237</i>
<i>Introduction and Discussion of Dr. Robert Freese.....</i>	<i>3-238</i>
<i>Introduction and Discussion of Dr. John Mallinson.....</i>	<i>3-240</i>
<i>Introduction and Discussion of Mr. Harriss Robinson.....</i>	<i>3-243</i>
<i>Introduction and Discussion of Ms. Carole Hogan.....</i>	<i>3-245</i>
<i>Introduction and Discussion of Dr. Reagan Moore.....</i>	<i>3-249</i>
<i>Introduction and Discussion of Mr. Bernard T. O'Lear.....</i>	<i>3-251</i>
<i>Introduction and Discussion of Dr. Kenneth Thibodeau.....</i>	<i>3-254</i>
<i>Introduction of Mr. Patric Savage.....</i>	<i>3-258</i>
<i>Introduction and Discussion of Mr. Adrian Abineri.....</i>	<i>3-259</i>
<i>Introduction and Discussion of Mr. Anthony L. Peterson.....</i>	<i>3-262</i>
<i>Introduction and Discussion of Mr. Glen Nakamoto.....</i>	<i>3-265</i>
<i>Introduction and Discussion of Mr. Michael Ruderman.....</i>	<i>3-269</i>
<i>Introduction and Discussion of Mr. Fred Rybczynski.....</i>	<i>3-270</i>
<i>Introduction and Discussion of Mr. Robert McLean.....</i>	<i>3-272</i>
<i>Introduction of Mr. Fred Zeiler.....</i>	<i>3-275</i>
<i>Introduction and Discussion of Dr. Bharat Bhushan.....</i>	<i>3-276</i>
<i>Introduction and Discussion of Mr. James "Newt" Perdue.....</i>	<i>3-277</i>
<i>Introduction and Discussion of Mr. Tom Gilbert.....</i>	<i>3-278</i>
<i>Introduction and Discussion of Mr. Ronald Blitstein.....</i>	<i>3-279</i>
<i>Introduction and Discussion of Mr. John Berbert.....</i>	<i>3-281</i>
<i>Introduction and Discussion of Mr. John Boyd.....</i>	<i>3-282</i>
<i>Introduction and Discussion of Mr. Fernando Podio.....</i>	<i>3-284</i>

MASS STORAGE CONFERENCE PROCEEDINGS

FIRST DAY - TUESDAY, JULY 23, 1991

On behalf of the Space Data and Computing Division and the NASA Goddard Space Flight Center, I would like to say it gives us great pleasure to welcome you here and how sincerely we feel in having this opportunity to bring together so many experts to participate in this important and timely conference on mass storage systems and technologies.

We at Goddard have a vested interest in on-line mass data storage systems and are anxious to learn from the experts gathered here what the field has to offer in solutions to some of the problems NASA faces in managing its large data systems.

The emerging role of mass storage media and their proper incorporation into the environments of NASA's computing centers and data centers is fundamental to the ultimate success of many of our scientific missions. In particular, it is a critical element to the success of one of the most ambitious projects NASA has ever undertaken or, for that matter, the science community has ever undertaken, namely the Earth Observing System, planned for launch towards the end of this decade.

But even before the launch of EOS, just managing the existing archives and the near-term major and moderate science discipline missions planned for the early to mid 1990's will provide a major challenge. For example, the Great Astronomical Observatory program, which includes the Hubble Space Telescope, the Gamma Ray Observatory, the Advanced X-Ray Facility, in addition to a host of moderate and complementary missions such as the Rosat X-Ray Mission, the Astro-D mission, and others will produce several terabytes of data a year.

Similarly, the International Solar Terrestrial Program and the Earth and Planetary Probes that are firmly scheduled for launch in the next two to three years, will consist of fleets of spacecraft which will produce tens to even hundreds of terabytes of data from imaging sensors and in-situ instrument measurements.

To archive and deliver this volume of data to scientists with reasonable response times constitutes a major technological challenge today. But NASA has added an additional requirement that the science community has requested, namely to develop and make available a much more enabling capability for remotely accessing and analyzing these data products.

NASA will attempt to bring into being a massive distributed collection of discipline specific data archives where all data are managed on-line and with transparent access to the users, so that scientists and other interested groups can have near-instantaneous access, from seconds to minutes, at worst, for browsing, querying, visualizing, and downloading data to their local workstations.

This presents very unique technical requirements to the developers of mass data storage systems that I believe have not yet been addressed in other applications. This is because NASA must deal with two types of data that are somewhat unique, namely very large and continuously evolving data sets from a finite number of instruments as well as data from very complex modeling simulations. I like to characterize these data types as nonrecoverable and recoverable data, respectively.

The unique aspects of these data types is not just their massive volumes that need to be stored on permanent, nondestructive, long-lasting media, but the need to accumulate these data over periods of time, which might grow to be in the climatic range itself, that is, data stored on the order of 25 to 100 years.

Already, scientists are beginning data assimilation experiments with conventional observations dating back to the 1950s and extending up to the present day. In addition, data must be maintained in such a way that complex computational simulation models, like numerical weather prediction models or solar cycle oscillation models, can time-continuously assimilate the entire period of multisensor observational data and multisets of computational experiments.

Finally, I would like to conclude with a remark or a prediction that I somehow remember reading somewhere, or perhaps paraphrasing and making it a part of my own psyche, that goes as follows: The 1970s was the decade of the introduction of supercomputers, which was responsible for the arrival of computational simulation for a new dimension to the scientific paradigm comparable to experimental science. The 1980s was the decade of the workstation and PCs, and that brought popular computing and data management and data visualization to the scientist's desktop. But the 1990s will be the decade for on-line interactive access to all the information knowledge that mankind has acquired. Such data systems will become the libraries of the future; and I hope that, by the end of this conference, we will begin to get a feel from the experts here whether this is indeed the direction of the 1990s. Thank you.

**Dr. Milton Halem, Chief
NASA Space Data and
Computing Division**

StorageTek.

N 9 3 - 1 3 1 3 2

Enterprise Storage Report for the 1990s

by Fred Moore
Corporate Vice President
Strategic Planning
Storage Technology Corporation

Enterprise Storage Report for the 1990s

Abstract

Data processing has become an increasingly vital function, if not the most vital function, in most businesses today. No longer only a mainframe domain, the data processing enterprise also includes the midrange and workstation platforms, either local or remote. This expanded view of the enterprise has encouraged more and more businesses to take a strategic, long-range view of information management rather than the short-term tactical approaches of the past. This paper will highlight some of the significant aspects of data storage in the enterprise for the 1990s.

ENVIRONMENT - 1990s

- **Storage and storage management are most pressing issues**
- **Networking and connectivity requirements increasing at all levels**
- **DASD subsystem reliability and availability continue to improve**
- **Fault tolerant DASD architecture emerging**
- **Strong acceptance of automated libraries**
- **New focus and mission for magnetic tape technologies**

As the 1990s begin, effective storage management remains possibly the most pressing issue. Poor device utilization and erratic performance are no longer accepted as normal conditions. The cost of ineffective storage management also has received considerable attention as storage costs now exceed the processor costs in mainframe environments.

The definition of the enterprise has moved quickly beyond the world of IBM mainframes to include other mainframes, midrange distributed processors and networks, local and wide area. The need to connect these processing platforms through standard network interfaces and provide access to common storage is increasing rapidly as most users now have mixed-vendor environments (Cray, DEC, IBM, Tandem, etc.) to manage.

The reliability of DASD subsystems continues to improve but even at 99.99 percent availability the only acceptable goal remains 100 percent availability. This trend has encouraged several companies to develop fault-tolerant DASD architectures. Fault-tolerant DASD subsystems provide continuous data availability in the event of any hardware component failure.

The successful introduction of automated tape systems such as StorageTek's 4400 Automated Cartridge System has led to widespread acceptance of automated storage. The 1980's view that library architectures were the least reliable component of a data center is now obsolete. The data processing industry has overcome preconceptions created by various mass storage and rail-type architectures of the past.

The highly successful launch of automated cartridge systems has given new life to tape data storage. Primarily used for low-activity backup, automation has allowed many new applications, not previously considered for tape, to become practical and cost-effective.

ENVIRONMENT - 1990s (Cont.)

- **Automated operations becoming a strategic goal for most large installations**
- **Accelerated new application growth in PC/workstation segment**
- **Outsourcing slowing**
- **Disk growth rates have moderated to 25-30% annually**
- **Image processing market slowly emerging**
- **IBM SYSPLEX re-focuses mainframe role**

Automated operations is quickly becoming a strategic goal of most large-scale data centers.

For the first time, there is now more storage outside the "glasshouse" mainframe environment than in it. This accelerated growth rate for storage away from the mainframe area will lead to system-managed storage structures, automated tape systems, multi-media libraries and fault-tolerant DASD for the midrange and desktop computing environments.

Outsourcing, a trend that gained considerable visibility in the late 1980s, has lost some of its appeal. Sourcing some or all computer operations, services and development to a source outside the enterprise is intended to save money. Though initial short-term financial gains are possible, many users now are viewing outsourcing as losing control of the most critical component of a business — the information processing function.

Mainframe DASD growth rates have moderated. In the early 1980s, DASD growth rates pushed a 60 percent annual increase in installed gigabytes. As the 1990s unfold, we note that most of the batch-to-online conversions are over. Secondly, data bases are now common, predominant in most organizations, eliminating redundant data. The third reason is that more users are beginning to make more effective use of storage management tools. Finally, many of the new applications that remain to be automated are emerging slowly such as image processing. Even at 25 percent compounded annual growth, the installed base of DASD capacity will double every three years.

It is believed that at the mainframe, midrange and workstation levels, image processing will be the dominant single driving factor for storage demand in the 1990s. This movement, however, has not evolved as quickly as projected due primarily to the lack of an effective enterprisewide image management architecture.

The role of the mainframe in the 1990s was clearly established by IBM's September 5, 1990, announcement of SYSPLEX. This announcement refocused the role of the mainframe in the enterprise as the central server and overseer of the networked enterprise. Mainframe architecture will continually drive many of the standards used for the entire enterprise.

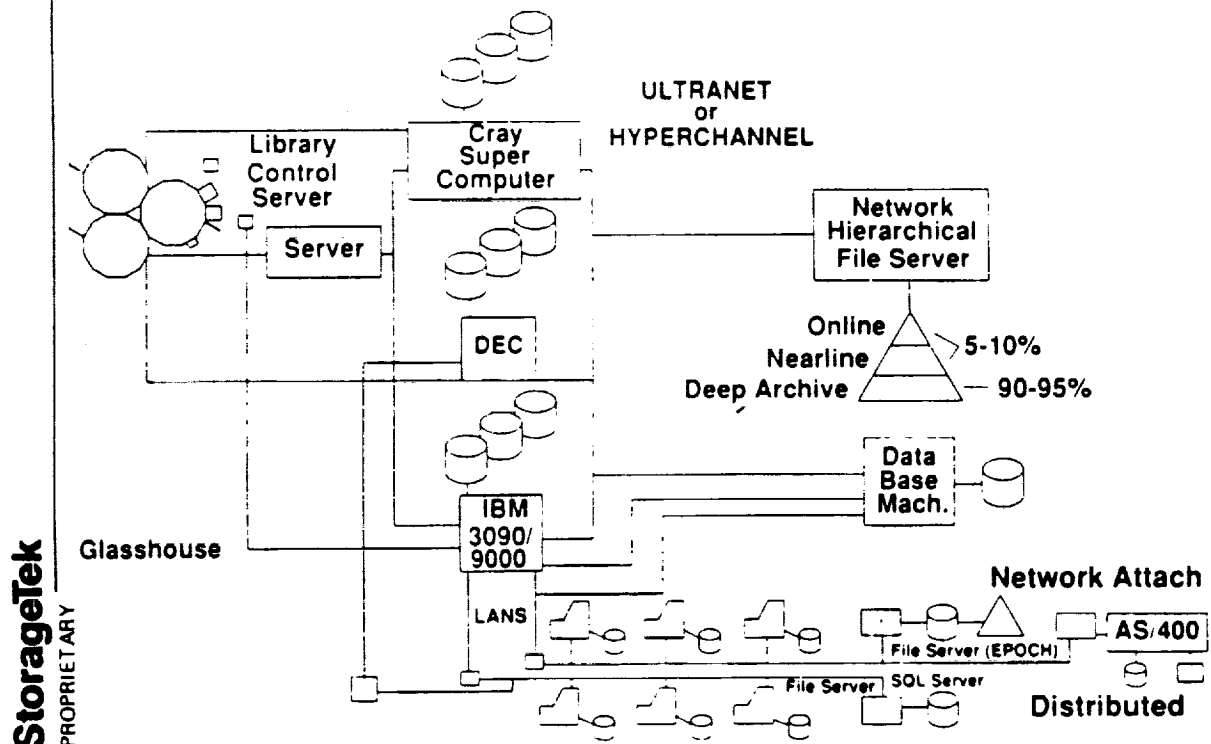
DEMAND TRENDS

	Processor Demand (MIPS)				Storage Demand (GB)
	1960-1969	1970-1979	1980-1989	1990-1999	1990-1999
Mainframes	15%	20%	40%	25%	25-30%
Midrange	--	35%	50%	40%	40%
Workstation/Desktop	--	--	150%	60%	55-65%

Let's take a look at some of the growth rates that we have seen in the last three decades and will see in the future. This chart projects CPU or processor demand. Notice that processor demand for mainframes in the 1990s is expected to be around 25 percent, measured in MIPS.

We observe 25 percent growth in the mainframe, 40 percent in the midrange, and the workstation growing overall at about 60 percent annually during the 1990s. Today MIPS demand corresponds almost one-to-one with storage growth. Storage management and the ability to access all data objects from all computing platforms will become both a requirement and a major architectural challenge of the 1990s. The vendor that can resolve this problem best will control the enterprise.

ENTERPRISE DATA STORAGE ENVIRONMENT



We have had the traditional glasshouse or mainframe view of data processing for a long time. That glasshouse today, dominated by MVS and VM environments, is beginning to share the spotlight with the rapidly emerging midrange and workstation/desktop processing environments. The enterprise now requires management of heterogeneous and complex environments. These three platforms are clearly and distinctly emerging as major areas of processing and storage for the 1990s.

Data processing now becomes distributed at nodes in the enterprise and the objective is to allow transparent access while maintaining the security, integrity and performance of the environment. The role of large systems in the 1990s will become one of management and control for the information enterprise.

We will see the migration of MIPS and storage, and the management issues that go with them, move from mainframe to midrange and desktop. We are not going to be able to limit our views of storage management to MVS and glasshouse and IBM-only for much longer. Storage management solutions must cross those architectural and communication boundaries.

Nearline is a registered trademark of Storage Technology Corp.

VIRTUAL STORAGE HIERARCHY

Average
Cost/MB Performance Capacity

\$4,000 80 ns 4 GB

\$1,560 75 μ s 16 GB

<\$900 1.5 ms nX GB

Central

Expanded
Storage

Auxiliary
Storage

- Contains operating systems, programs and pages

- Non-shared data, required for HSP1
- Can contain HSP2, data spaces and pages

- Paging devices, SSD and DASD
- Contains HSP2 and data spaces in page datasets

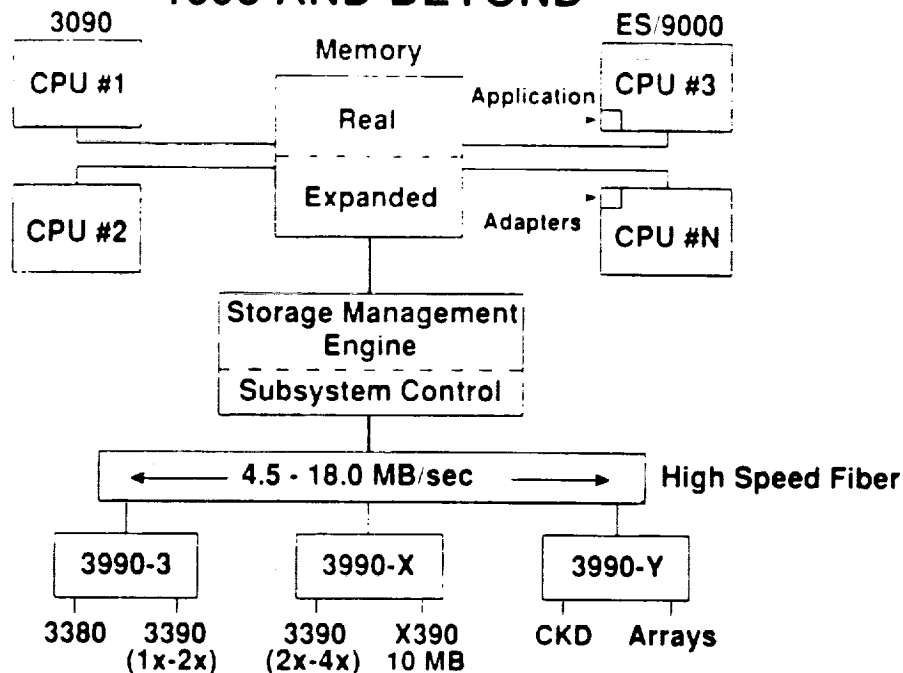
StorageTek
PROPRIETARY

The virtual storage hierarchy of the 1990s will be exploited by ESA/390 architectures and consist of three levels of storage. Central memory, at about \$4,000 per megabyte, has an architectural limit of 4 gigabytes and an announced 1 gigabyte limit. Expanded storage has a 75-microsecond access time. The limitation in present ESA/390 architecture is 16 gigabytes of addressable expanded storage, though the announced limit is 8 gigabytes.

Solid-state products are now considerably less than \$900 per megabyte and the cost per megabyte is declining quickly. In 1979 when the first SSD was introduced by Storage Technology Corp., the original price was \$8,800 per megabyte. We have seen over a 90-percent reduction in pricing on solid-state technology in the 1980 decade. The virtual storage hierarchy contains three levels including auxiliary or paging storage. Careful use of all three technology levels offers the most cost-effective solution to managing the virtual storage hierarchy. It is normally not cost effective for most users to place all performance-critical data in expanded and central storage.

The ES/9000 processor series now permits migrated pages to move directly from expanded storage to the channel subsystem (auxiliary storage paging) improving the synergy and performance between both levels of the virtual storage hierarchy.

LARGE SYSTEMS ARCHITECTURE 1993 AND BEYOND



StorageTek
PROPRIETARY

IBM's recent SYSPLEX announcement clearly refocused the role of the mainframe in the 1990s. Let's examine a likely scenario for the 1993 time frame and beyond. This has often been referred to as the post-Summit or Future Systems (FS) architecture. It is expected that up to a limit of 16 (CPU #N) ESA-based processors evolving within a SYSPLEX could be connected in the manner shown.

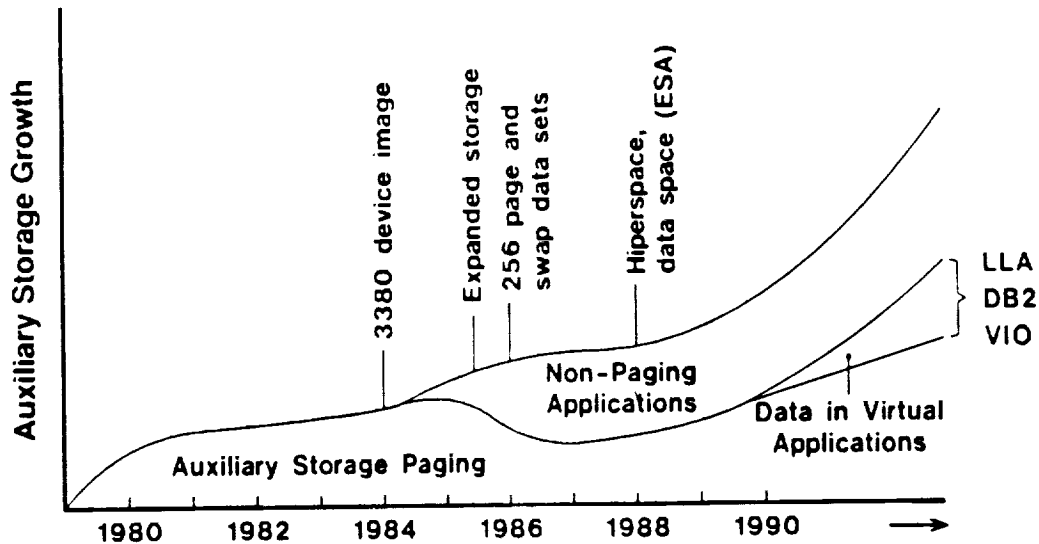
The continued roll-out of this architecture will include a shared expanded storage capability and application-specific adapters implemented via software, licensed internal code and hardware in varying amounts. Hardware assists for DFSORT announced in the September 5, 1990, IBM announcement are using this concept.

A storage management engine or I/O processor will be a new concept used to off-load from the host processor many of the I/O functions such as parts of IOS (I/O Supervisor), VSAM and DF (Data Facility) functions. The storage management engine will attach peripheral devices as we know them today (SSD, DASD, tape, printers and terminals) via the channel subsystem. Attachment of ESCON (Enterprise System Connectivity) serial fiber channels will be preferred though parallel bus and tag channels will need to attach via ESCON converters. The point-to-point limit of ESCON channel transfer rates is 18 MB/sec.

ESCON is a trademark of IBM Corporation

DIRECTION OF AUXILIARY STORAGE RAM BASED ARCHITECTURES

StorageTek
PROPRIETARY



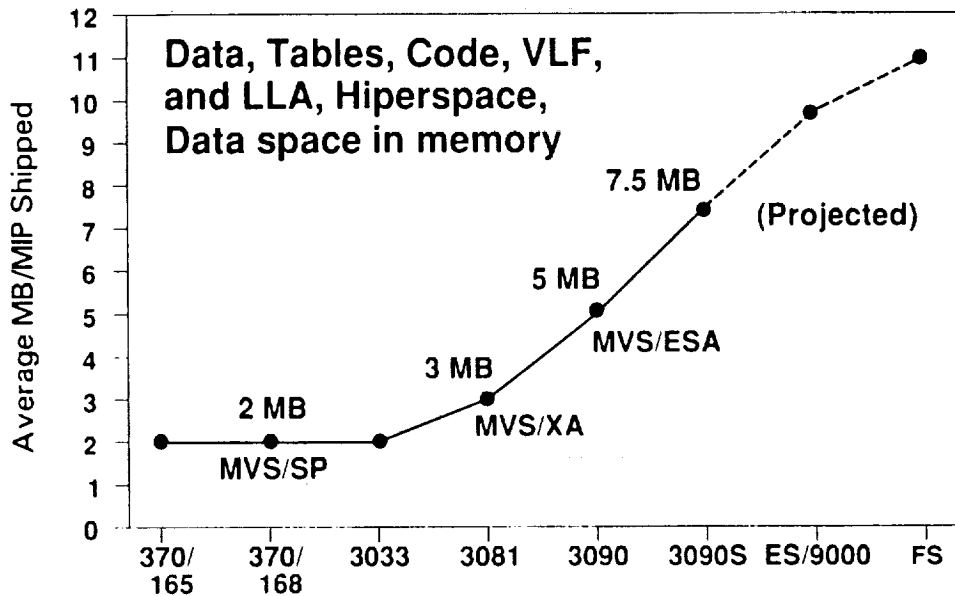
By examining RAM-based solid-state products, you will notice their use from 1979 through 1984 was exclusively for paging. When expanded storage appeared in 1985, solid-state devices became viewed as a high-performance disk, and non-paging data such as load libraries, catalogs and indexes were placed on solid-state devices. The 1990s will see the MVS/ESA and VM/ESA hiperspace and data space applications drive up data in virtual requirements and force users to seriously consider using SSD as a cost-effective complement to real and expanded storage. ESA/390 drivers of virtual storage consumption, called methods of I/O avoidance by some, will include linear VSAM, Virtual Lookaside Facility (VLF), hiperspace catalog, hiperspace buffers and DB2. MVS/XA systems previously using 400 to 500 megabytes of auxiliary storage will soon identify requirements exceeding 1 gigabyte or more after migrating to ESA.

DRAM MEMORY DENSITY PROJECTIONS

Technology (in microns)	DRAM Technology	Development Start	Introduction Date	Peak Output
1.8	256 Kbit	1977	1982-83	1988
1.2	1 Mbit	1980	1985-86	1991
0.8	4 Mbit	1983	1988-89	1994
0.6	16 Mbit	1986	1991-92	1997
0.36	64 Mbit	1989	1994-95	2000
0.25	256 Mbit	1992	1997-98	2003
0.15	1 Gbit	1995	2001-01	2006

Unlike rotating DASD, RAM-based architectures command a very price-elastic market. If the price decreases, the demand increases. You cannot necessarily stimulate the demand for DASD or tape by changing the price. The industry-standard DRAM chip has moved from 1 megabit to 4 megabits. Note that the 4-megabit chip has been under development since 1983. As DRAM densities increase, price decreases along with the physical space required to store information. Thus much higher capacity DRAM storage devices will appear occupying smaller footprints. This trend should continue until the point where DRAM-based storage devices will occupy a large portion of the storage hierarchy currently belonging to rotating and cached DASD.

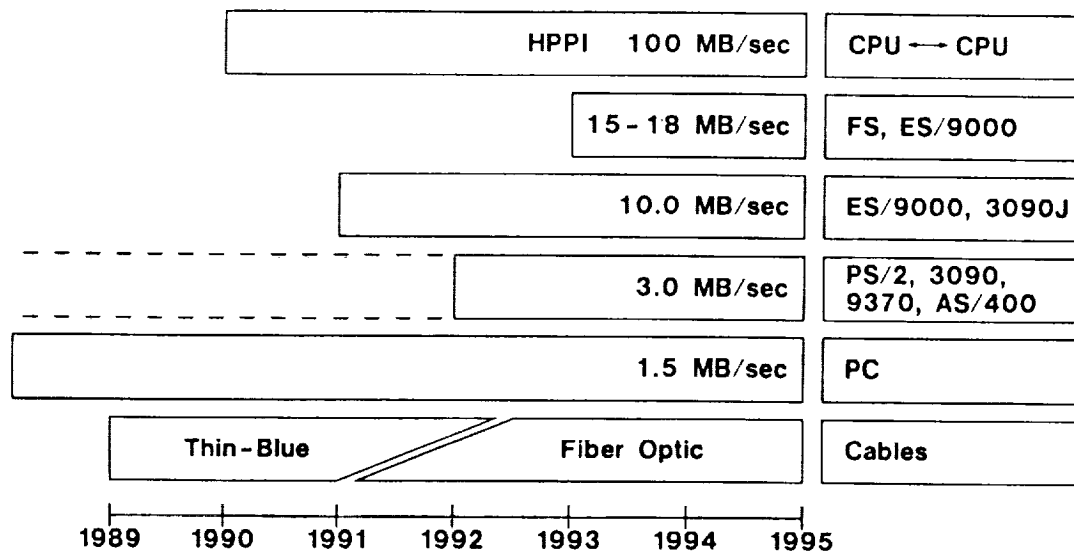
MEMORY/MIPS SHIPPED



Historical tracking of the ratio of memory installed per MIPS installed indicates the trend increasing sharply with MVS/XA (31-bit addressing) and MVS/ESA, effectively 44-bit addressing. The MVS/ESA capability to place data, and program load libraries and other objects into memory, will continue to drive the ratio upward, requiring more and larger RAM-based storage solutions. The announcement of VM/ESA and DOS/VSE/ESA implementing hiperspace and data space concepts into these operating systems will further encourage virtual storage growth. The growth rate grows sharply until shared expanded storage arrives late in the Summit (i.e., future systems) then moderates slightly.

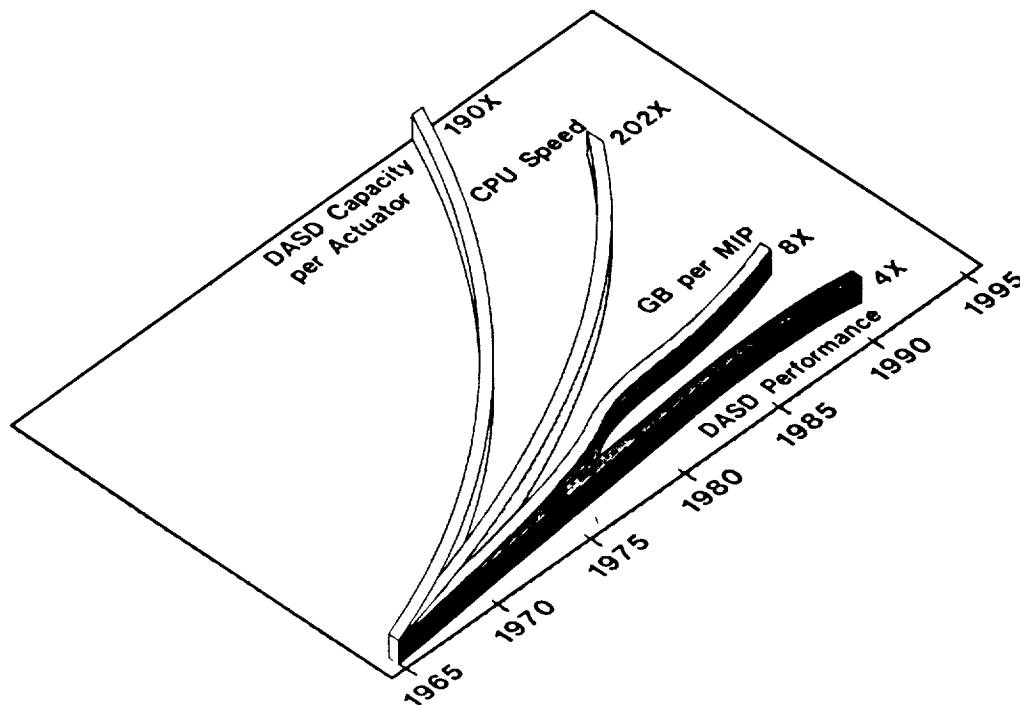
CHANNEL TRANSFER RATE GROWTH PROJECTION

StorageTek
PROPRIETARY



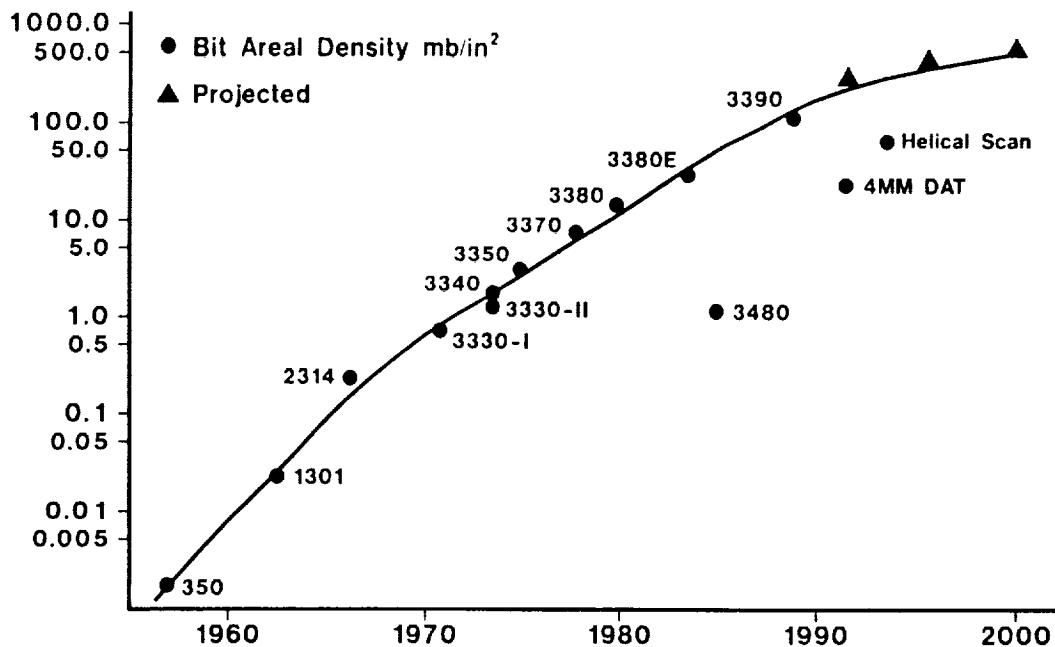
The data transfer rate capabilities at all levels of computing will increase much faster in the first half of the 1990s than they did in the previous 15 years. ESCON channels now offer up to 10 MB/sec data transfer rates. Device exploitation of ESCON channels at 10 MB/sec will come more slowly. The 3990-3 DASD Storage Control will be the first device to exploit ESCON channels at ESCON speeds. Up to 18 MB/sec on the FS series is likely by 1995. The gray and thin blue cables will begin to give way to serial fiber channels providing increased distance (up to 9 km initially) in the early 1990s. The HPPI (High Performance Parallel Interface) channels used by large-scale CPUs for scientific processing will offer attachment of specialized storage devices in the mid-1990s, likely RAM-based, providing high-performance solid-state storage arrays.

LARGE SYSTEMS TRENDS



In the last 25 years, the processing power of large computers has increased at a much faster rate than the performance capabilities of the I/O subsystem. Since the introduction of System/360 in 1965, we have seen the capacity of a disk actuator increase 190 times. Processor performance has increased over 200 times, but the performance of a disk actuator has improved only 4 times. This divergence of processor speed and the I/O subsystem performance has been the subject of considerable interest, particularly in the 1980s. During this time, we have seen the introduction of a number of technology developments to help bridge the gap. These enhancements include solid-state disk, cached control units, dual port, quad port, actuator level buffers, tape buffers and expanded storage. Despite these advances, the performance gap between processors and I/O subsystems continues to diverge. More solutions will emerge, predominately based on DRAM technology, to place data closer to the processor and remove the performance delays of mechanical devices. The ES/9000 processor announcement by IBM is a good example of this continually diverging trend — processor speed (MIPS) nearly doubled while the speed of the storage subsystem remained unchanged.

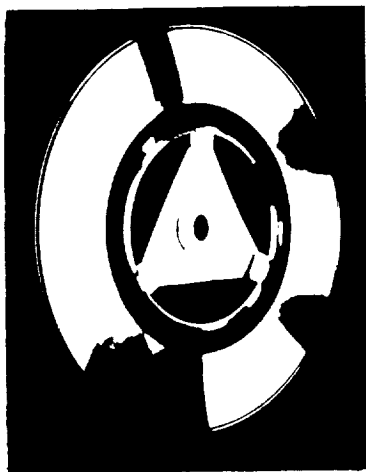
TRENDS - AREAL DENSITY



Magnetic recording technology has seen few of the anticipated limitations in the areal density (megabits per square inch) that were expected to occur in the early 1980s. Areal density increases in DASD, once not expected to exceed 30 megabits per square inch, are now over 60 megabits per square inch in 3390-type disk drives and are expected to increase to over 500 megabits per square inch by the end of the 1990s. Laboratory developments of 1 gigabit per square inch have been demonstrated. As the areal density of magnetic recording continues to increase rapidly, the future role of optical disks in the large systems storage hierarchy becomes more questionable.

The helical scan tape format, recording data tracks on a tranverse rather than parallel to the edge of the tape, offers areal densities in the range of 30-50 million bits per square inch and very high data transfer rates. Helical scan technology should enter the hierarchy at the deep archive level and co-exist with 3480 chromium dioxide (Cro2) tape format throughout the remainder of the 1990 decade.

OPTICAL



WHEN

MAGNETIC/OPTICAL STORAGE

5 1/4" OPTICAL PARAMETRICS vs DASD

OPTICAL HDA	1990	1995	2000
Capacity (GB)	1.0	1.5	4.0
Transfer rate (MB/sec)	0.7	1.2	2.5
Cost (\$/MB) (OEM)	3-5	1.30	0.50

DASD HDA	1990	1995	2000
Capacity (GB)	1.5	2.5	5.0
Transfer rate (MB/sec)	1.9	4.0	6.0
Cost (\$/MB) (OEM)	2.0	0.60	0.20

Optical disk storage has been hailed as the low-cost mass-storage technology of choice for years. In reality, optical storage has yet to fulfill expectations. Issues such as standardization, throughput, data transfer rates and uncertainty of shelf life (data retention) remain. The areal density of magnetic disk storage is increasing rapidly, while optical disk areal density has remained relatively the same for the past six years. The write-once, read many times (WORM) drives are common in the midrange and desktop markets, but will struggle to find a mainframe niche. Magneto-optic or erasable optical disks offer the large systems market the most benefit, but may face a stiff challenge from large-capacity DASD array storage solutions for the online, large-capacity storage market and advanced automatic cartridge systems for even larger capacity and less costly deep archive storage. Unlike WORM optical, 5.25" magneto optical has the support of formal standards by the International Standards Organization.

TAPE LIBRARIES MAINSTREAM APPLICATIONS

StorageTek

CURRENT APPLICATIONS

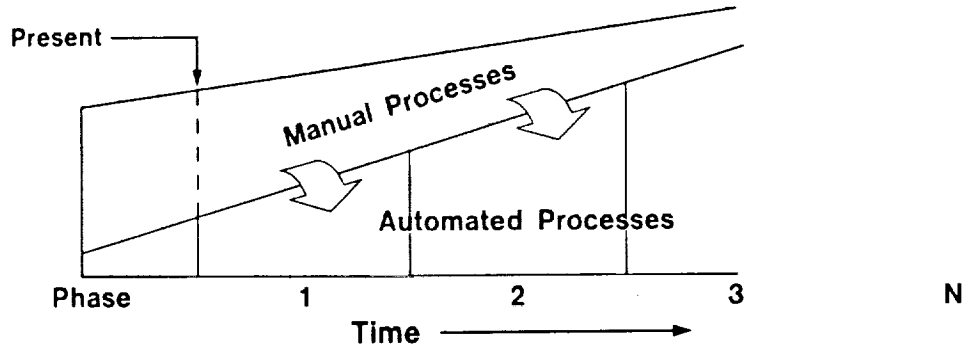
- | | |
|---|--|
| <ul style="list-style-type: none"> • Tape Management • DASD Management <ul style="list-style-type: none"> - SMS - DF/HSM, DMS/OS - DASD savings • Job scheduling and rerun <ul style="list-style-type: none"> - Improved batch performance • Software development | <ul style="list-style-type: none"> • Automated recovery <ul style="list-style-type: none"> - Online data bases - Mission-critical data • Report management/paper/fiche • Electronic archive <ul style="list-style-type: none"> - Campus - Remote vault • Automated operations <ul style="list-style-type: none"> - Unattended → Lights out • High speed search applications |
|---|--|

ADVANCED APPLICATIONS

- | | |
|---|---|
| <ul style="list-style-type: none"> • Anticipatory staging <ul style="list-style-type: none"> - Random access data - Data set scheduling | <ul style="list-style-type: none"> • Network storage • Deep archive with helical formats • Scientific data |
|---|---|

Automation has opened many new horizons for data storage. Far beyond simply automating what currently exists on tape today, automation has become a primary ingredient in cost-effective storage management and is enabling the promise of systems managed storage to be fulfilled. Several new applications listed above have become areas of opportunity providing cost savings and improved quality of operations beyond prior capabilities. The concept of electronic archiving has been given an increased focus with the announcement of ESCON channel architecture complementing traditional channel extension methods. A form of image storage, report management, has been enhanced with a number of software products that allow computer output microfiche and printed data to be stored on a tape library, viewed at a computer terminal, and printed or sent to fiche only if needed. This new area of library exploitation greatly reduces distribution, copy and filing costs while improving the security aspects associated with printed storage. Applications with much promise for automated tape libraries include anticipatory staging of data and deep archive storage for long term-data storage.

AUTOMATED OPERATIONS



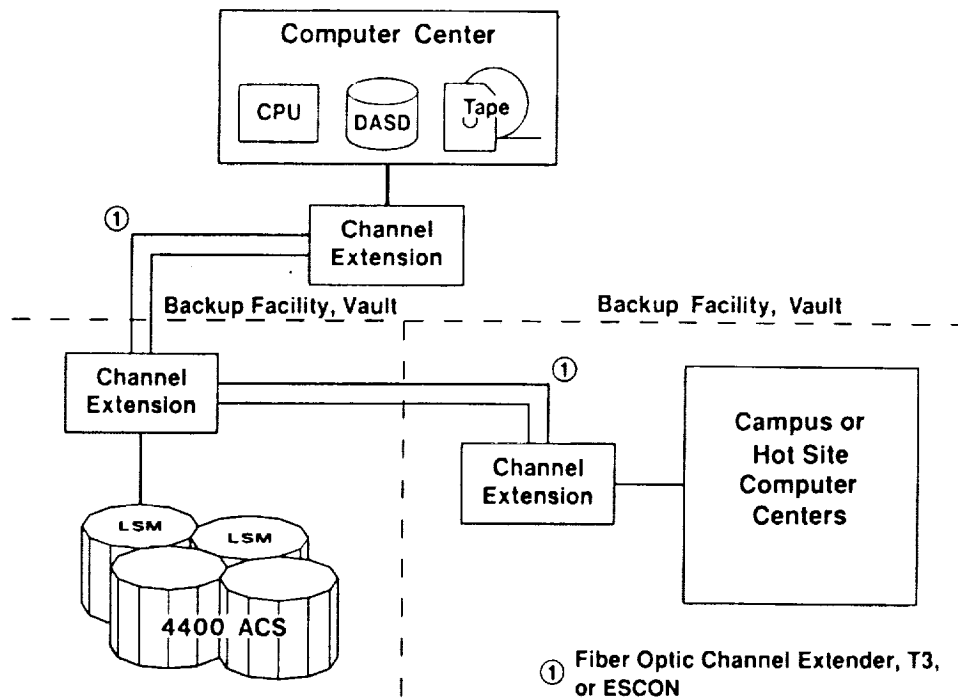
StorageTek
PROPRIETARY

Console Message	Operator	Suppress Auto Reply	Open Architecture
Recovery	Manual	Auto Restart	D/B Recovery Appl. Recovery
SMS	Storage Pools	Space Mgmt. DF/SMS	Performance Management
TMS	Physical Placement	Scratch Retention	ACS 4400
Output Distribution	Burster Trimmer	Roll Feed Bar Code	Online Viewing Selective Print

As automated operations becomes a strategic goal for many data processing users, solutions are appearing which are making companies more competitive, more productive and more profitable. Automated operations is usually fully implemented in stages and will evolve to include expert systems solutions to resolve some of the complex, enterprisewide information management issues. The primary reason for automated operations is improved quality of the data processing organization.

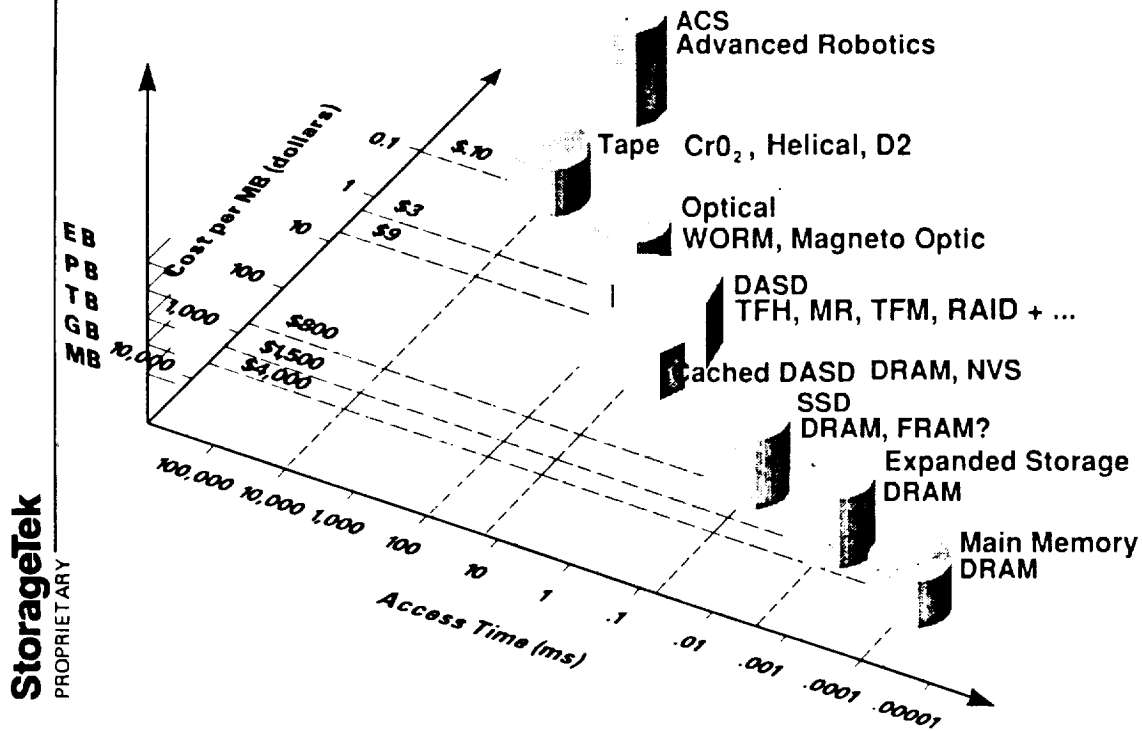
In addition to automated operations, business resumption or disaster recovery planning has become a strategic goal for many users. Workshops on these areas of advanced data center operations are available on a worldwide basis from Storage Technology Corp.

ELECTRONIC ARCHIVING



The rapid acceptance of library storage products such as StorageTek's 4400 Automated Cartridge System (ACS) has provided a means for computer users to archive data electronically to a secure, remote location such as a vault or warehouse. Today, the use of fiber optic channel extensions provides 3 megabyte per second device attachment. Products such as the 4400 ACS can be located at distances well beyond the four walls of the computer center by using fiber optic channel extenders, T3, or ESCON communication lines. The "data vault" provides backup of critical data in a safe location and also can link into a hot site or campus computer facility for a quick recovery in case of a disaster. This trend will expand in the 1990s as the value of information to the corporate enterprise becomes increasingly more important.

Progress of Storage Technologies



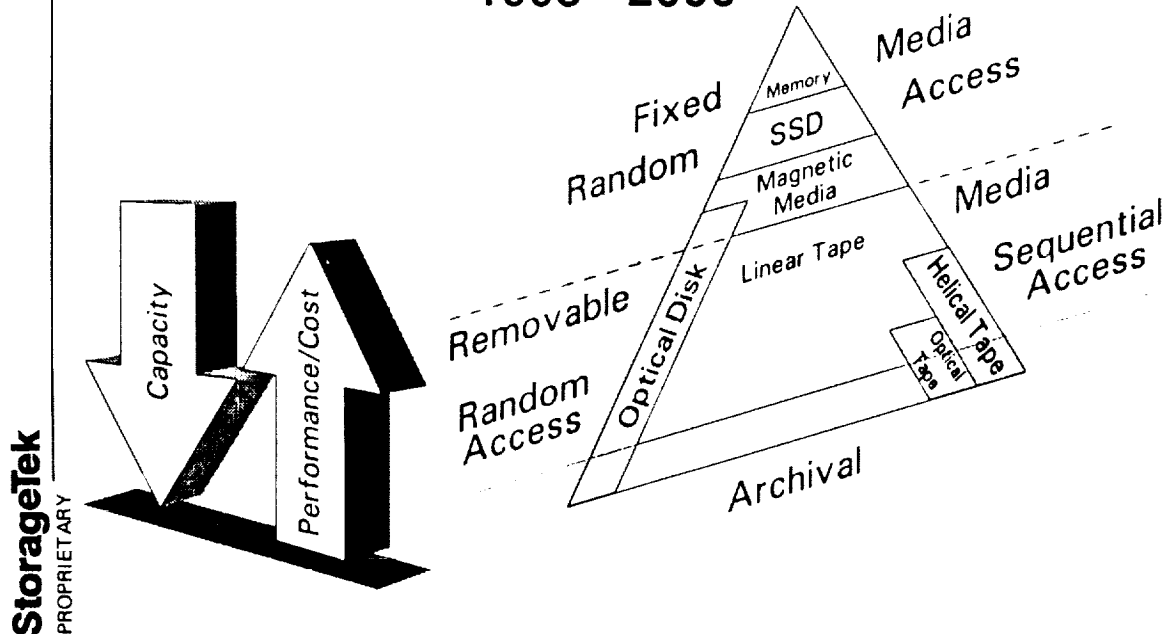
Let's take a look also at a few of the technologies that have merged to help resolve some of the challenges in the 1990s. The 1980s was the the decade of technology. The 1990s will be the decade of how we exploit that technology effectively. In the 1980 timeframe, we clearly remember vendor and customer discussions regarding DASD. Issues centered on such things as the diameter of the disk platter. How thick is the platter lubrication? How high does the read/write head fly? The answers to those questions sometimes influenced buying decisions. Today in DASD acquisitions the issues are gigabytes per square foot; I/Os per second; availability; cost per gigabyte. The size of the platter really doesn't have to make a difference, but gigabytes per square foot should.

Notice that main memory presently is priced at about \$4,000 per megabyte with an access time measured in nanoseconds. Expanded storage has a 75 microsecond access time and is priced at about \$1,500 per megabyte. There remains a major performance and price gap between the memory technologies (DRAM) and moving or rotating technologies.

Optical storage and tape provide two interesting comparisons. The areal density of optical storage has witnessed insignificant improvement in the last six years. During this time, magnetic storage has significantly increased in areal density. Interestingly, the IBM Image Plus system, using a write once optical storage device, is priced around \$2 to \$3 per megabyte. Tape libraries, including compression/compaction, may realize costs as low as 10 cents per megabyte.

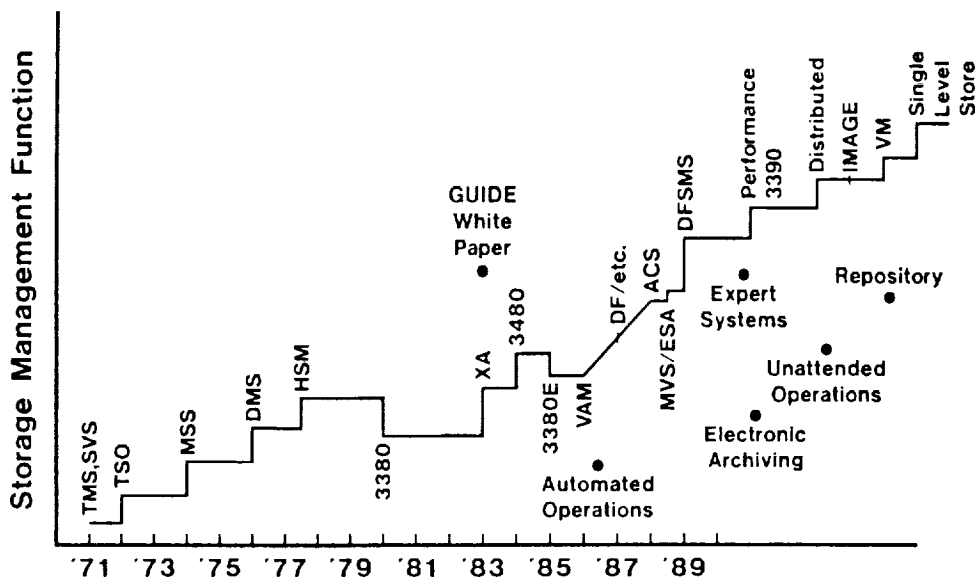
New technologies will evolve to fill this gap such as ferro-electric RAM (FRAM) devices. These are non-volatile RAMs and still under development though expected to be affordable and commercially available in the 1994 time frame.

STORAGE HIERARCHY 1995 - 2000



The hierarchy of storage technologies, not devices, in the last half of the 1990s is shown. This hierarchy may be broken down into fixed-media and removable-media segments. Fixed-media storage will consist of RAM and rotating DASD. Optical disk, once viewed as the heir apparent to the removable-media segment, has given way to automated tape systems that are faster and less costly. Further advances in the capacity of 3480 cartridge capacity to 1 gigabyte levels and the increased usage of helical scan formats make magnetic tape (along with magnetic disk) the key technologies of the 1990s.

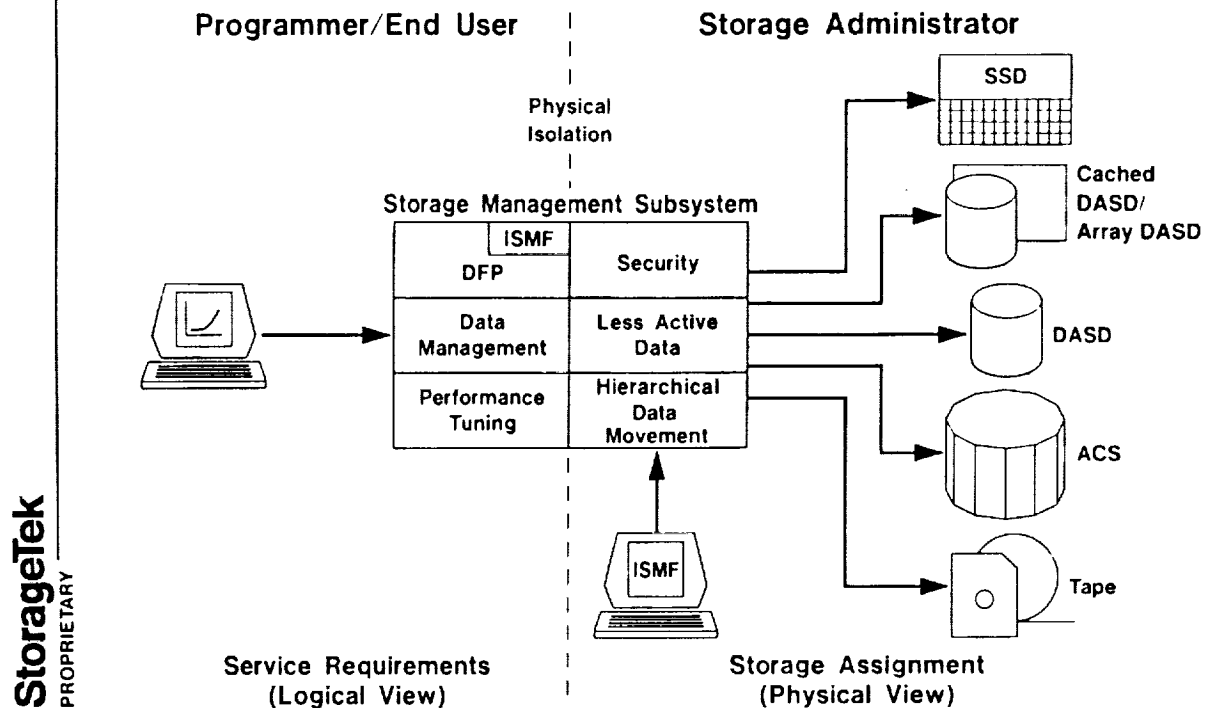
TRENDS - STORAGE MANAGEMENT EVOLUTION



The total cost of storage devices now accounts for more than one half of the total hardware expenses of the typical large data center today. In the early 1970s, storage management meant tape management. With the introduction of the DMS/OS and HSM storage management products in the late 1970s, storage management expanded its scope to include space management for disks. Since that time, storage management had been relegated to improving various facets of space management until the announcement of DFSMS in February 1988. This platform should gradually evolve to include dynamic performance tuning, storage management for distributed processing nodes, networks, workstations, a DFSMS equivalent for VM and development of a repository to identify objects across all computing platforms in the corporate enterprise.

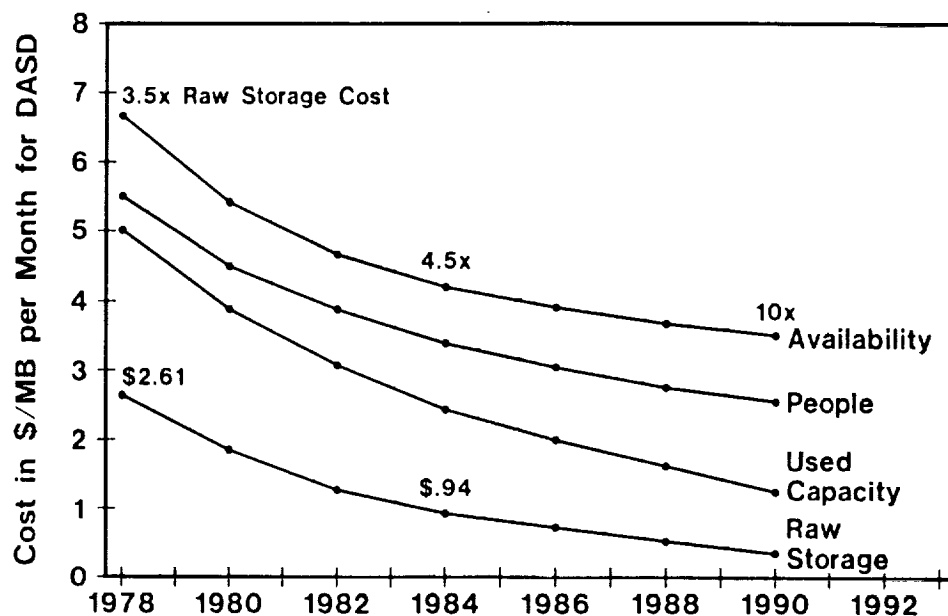
DFSMS is a trademark of IBM Corporation

SYSTEMS MANAGED STORAGE



Systems managed storage is a concept that allows an operating system to perform many of the human-intensive processes involved with space management, performance tuning, availability management and true hierarchical storage management supporting all tiers of storage. Though in its infancy, systems managed storage must efficiently evolve these processes to provide the platform for single-level storage in the last half of the 1990s and achieve an environment that allows "true systems managed storage." DFSMS is one of several products that make up systems managed storage. Presently the DFSMS product provides no performance tuning capabilities or movement of data vertically throughout the storage hierarchy to optimize performance or space management.

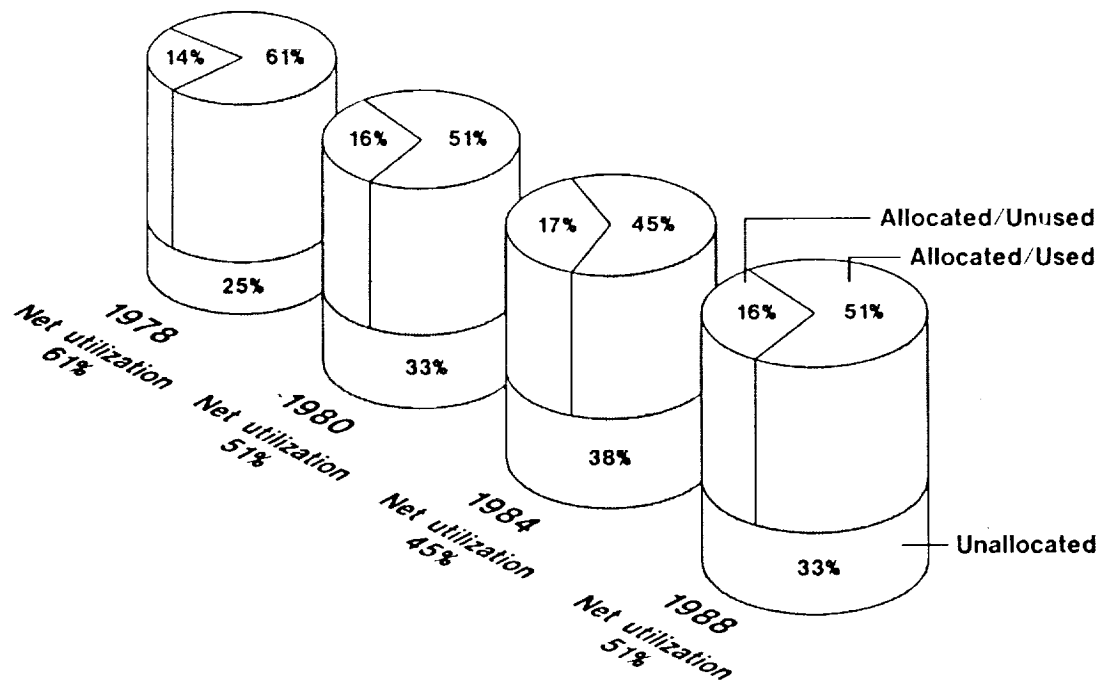
COST OF USABLE, MANAGED STORAGE



Source: IBM Systems Journal, 1989

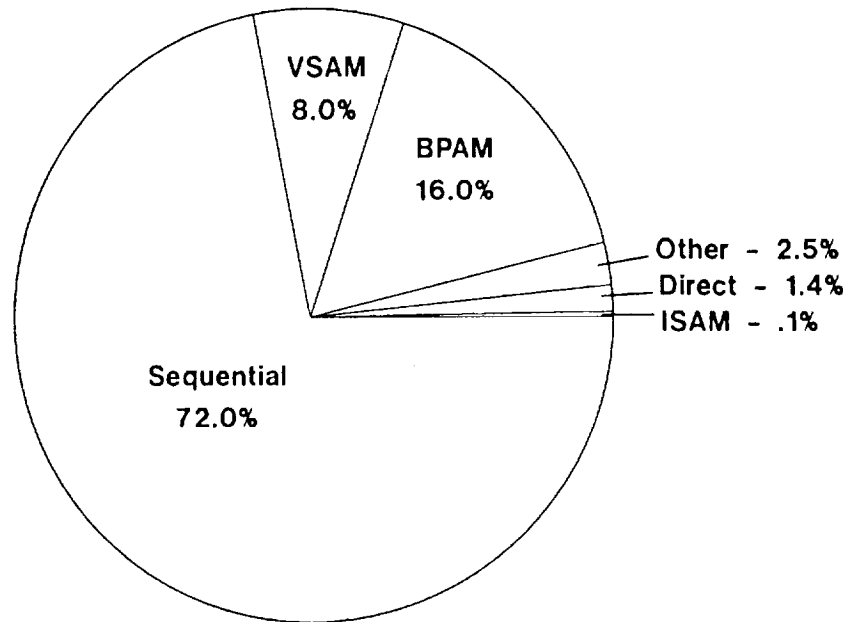
The total cost associated with managing storage has declined since 1978 on a per-megabyte basis. The cost of raw (live) data, plus the costs of unused capacity, people costs to manage the storage system and additional costs of availability such as backing up data, have increased several times since 1978 on a per-megabyte basis. The cost of managing DASD, effectively or ineffectively, has become a primary concern in very large (terabyte plus) data centers and clearly is necessitating the movement toward much improved storage management facilities. It is estimated that the total cost of managing disk storage in the 1990s will be as much as 10 times greater than the cost of actual data stored on DASD. New storage solutions, such as advanced fault tolerant DASD architectures, will dramatically improve these trends.

DASD USAGE TRENDS



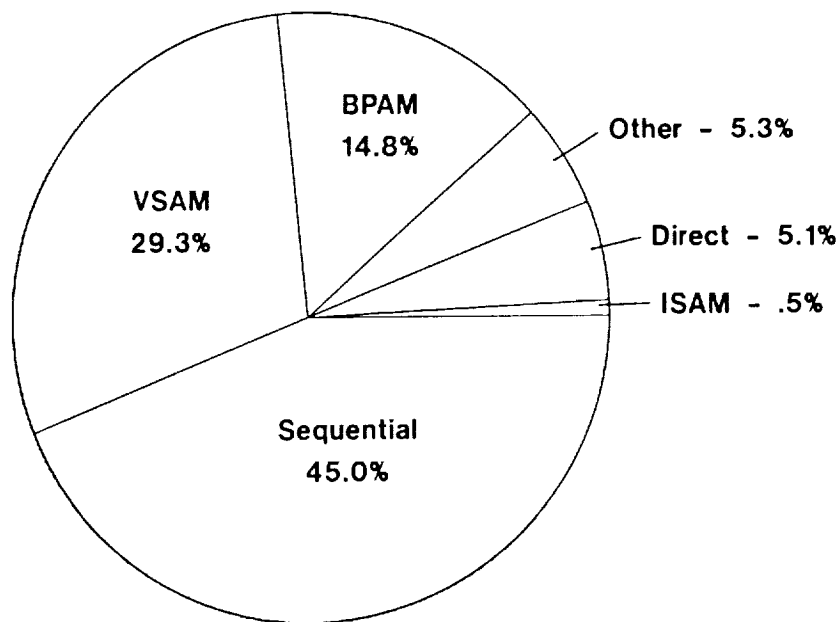
Formal surveys on DASD usage have been conducted since 1978. The net utilization, or the amount of real data on the average device decreased from 61 percent in the 1978 survey to a low of 45 percent in 1984. The 1988 survey indicated an overall increase in net utilization to 51 percent. This survey included single-, double- and triple-capacity 3380-type devices. The single- and double-capacity devices actually increased in utilization; however, the triple-capacity devices continued the downward trend. Utilization of the single- and double-capacity devices increased largely due to the higher percentage of caching permitting increased space allocation on cached DASD. Space utilization figures for 3390-class devices are not presently available though the lack of widespread DFSMS usage to optimize data allocation on 3390-type DASD may initially inhibit more effective utilization.

PERCENT OF DATASETS BY DATASET ORGANIZATION



As a follow-up to the previous chart on DASD space allocation, the percent of DASD data sets by data set organization reveals a correspondingly high percentage of sequential data sets. VSAM and SAM-E (sequential) data set organizations are strategic while partitioned data sets will fold into VSAM format as ESA/390 evolves. Other data sets such as graphics access methods, direct access files and even ISAM will exist as they are today. This profile again reflects the results of extensive tape-to-disk migration activities in the 1980s.

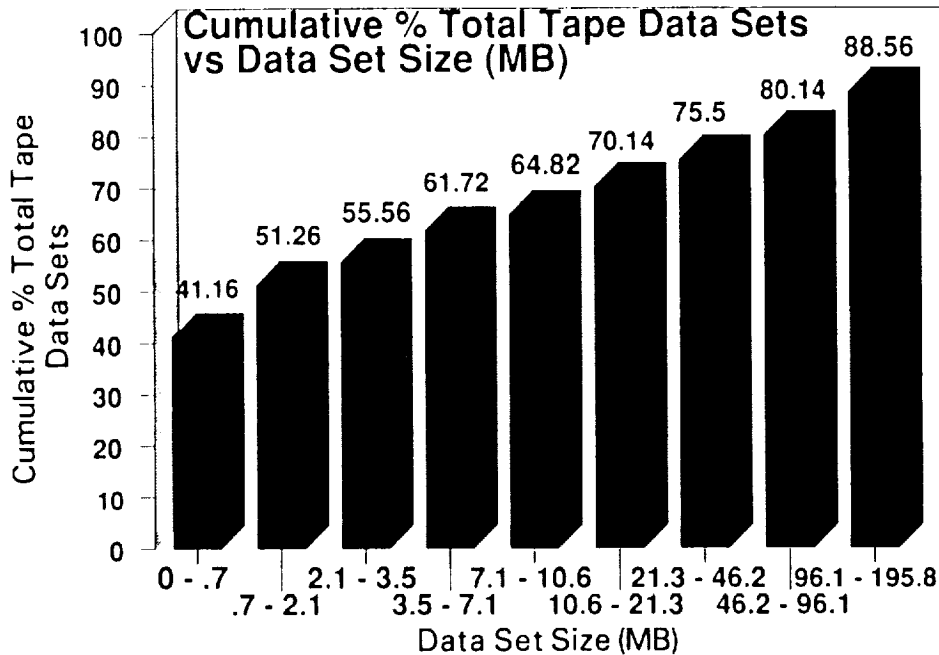
PERCENT OF SPACE BY DATASET ORGANIZATION



Storage management has increasingly focused on DASD in the last few years, primarily due to the criticality of data residing on DASD. At the end of 1988, sequential data had grown to include 45 percent of allocated disk space. This was primarily a result of the large number of tape-data-set-to-disk conversions in the early 1980s, occurring for the lack of any successful automated tape library system available to mainframe users. Tape data sets requiring rather quick or frequent access could not often withstand erratic human tape mount times.

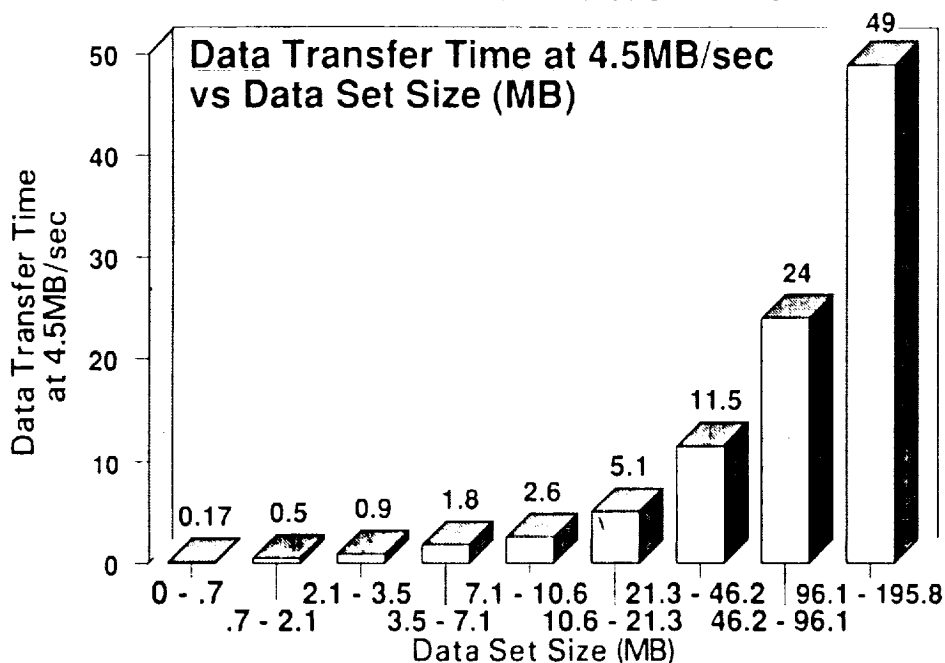
This DASD profile, as a result of the 1980s strategy of "put it all on disk," has become a significant cost savings target for automated cartridge systems in the 1990s.

TAPE COMPRESSION/COMPACTION TAPE DATA SET PROFILES



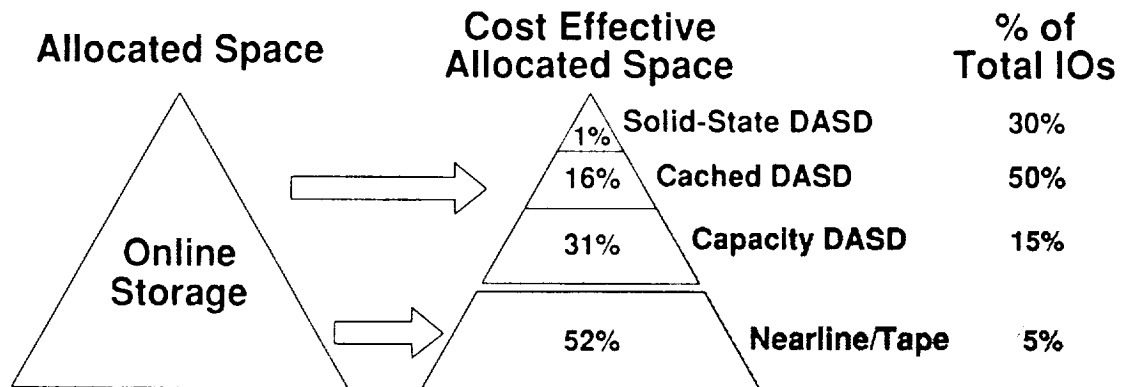
A survey of tape data set sizes indicates that the vast majority of tape data sets are very small. 70.14 percent of the tape data sets in the survey are under 21.3 megabytes in size. Even with the tape-to-disk conversions of the 1980s, many small tape data sets remained on tape and became obvious targets for automation. While 11.44 percent of the tape data sets occupy over one 200MB 3480 tape cartridge, the majority of these data sets are backup applications using processor data-compression functions.

TAPE COMPRESSION/COMPACTION TAPE DATA SET PROFILES



The tape data set size survey is equated to data transfer time for each classification of data set on a 4.5 MB/sec channel using bus and tag or even ESCON. The 21.3 MB tape data set takes approximately 5.1 seconds to process once mounted. Assuming a 2-to-1 data-compression factor, a total time of 2.55 seconds would be used to transfer the 21.3 MB data set. This minimal gain from compression/compaction is often lost by the variability in manual tape mounts resulting in minimal, if any, throughput benefit for 70 percent of the tape data sets. By the early 1990s, nearly everything written to tape will be compressed. Like 7-track, 9-track, NRZI and GCR in the past, tape data compression will be another format for tape data recording.

COST-EFFECTIVE STORAGE PROFILES



1980s

- Technology decisions
- Get the right data in the right place

1990s

- Business decisions
- Get the right data in the right place at the right time

StorageTek
PROPRIETARY

Using advanced data storage placement methodologies such as CPIO (Cost Per I/O) analysis, it is possible to determine the most cost-effective location in the storage hierarchy. Using the size, performance and storage cost per I/O to determine optimal data set placement, studies indicate that online storage users today may not be cost effectively utilizing online storage. Typically, 1 percent of the space allocated in online storage generates 30 percent of the total online IOs. This small but highly active group of data sets is most cost effectively located on solid-state disk. At the other end of the spectrum, a little over half of the data allocated on DASD today is more cost effective on automated cartridge systems or manual tape. Storage management in the 1980s stressed getting the right data in the right place; storage management for the 1990s will stress getting the right data in the right place at the right time.

CPIO is a proprietary software tool from Storage Technology Corp.

notes

OPTICAL DISKS BECOME ERASABLE

Dr. Robert P. Freese

President, Alpatronix, Inc.

Research Triangle Park, NC

Discussion

- * Optical Recording**
- * How Does it Work?**
- * Why All the Fuss?**
- * State of the Industry**
- * Sample Applications**
- * Future Directions**

OPTICAL DISK AND TAPE TECHNOLOGY

Dr. Robert Freese

DR. FREESE: Thank you. Good morning. Does everyone want to stand up and take a quick break?

(Participants stand up and stretch)

DR. FREESE: Okay. Now, it is 10:00 and let's begin. Just a quick fix.

What I would like to do today is spend a few minutes talking about and giving you an overview of the status of the optical recording industry.

(Showing of slides)

DR. FREESE: I have titled this particular overhead optical disk "become erasable." "Rewritable" is a technical term.

(First slide)

In today's presentation, I would really like to take an end user perspective, not so much of a technology perspective; and I would also like to take a more practical approach to the status of the industry, which is to say we have all sat through presentations and talks like this over the last decade and heard about all the wonderful things coming up.

But I would really like to take a more practical approach: What can I buy now? How do I use it? Why do I use it? And what am I going to see in the immediate future?

I would like to focus first of all just on the status of the optical recording industry, review how it works, talk about why all the interest and why all the fuss, go through a few sample applications and a few future directions.

(Change of slide)

DR. FREESE: Of course, everybody always asks me. There are three types of optical recording systems: the so-called read-only, the so-called write once, and the erasable.

(Change of slide)

DR. FREESE: Read onlys, just like the name implies: You can read the disks, but you can't record on them, and you can't write on them. The first implementations of this were like video disks; and in fact, I believe that you at NASA have used as data distribution mechanisms in the laser disk format.

(Change of slide)

More recently, CD ROM and CDI have been introduced in the smaller format and are starting to see widespread acceptance.

(Change of slide)

DR. FREESE: Probably the most questions I get are relative to WORM, or the write-once optical disks. WORM stands for "write once/read many times." These were introduced in 1982, 1983, and 1984--this type of time frame. Their marketplace is extremely small; we have seen that. It is really a precursor to the rewritable marketplace.

PRECEDING PAGE BLANK NOT FILMED

Optical Recording Systems

Three Types:

- * Read Only
- * Write Once (WORM)
- * Erasable

Read Only Optical Disks

Introduced: 1979

Function: Read (only)

Markets: Music, Publishing

Application: Compact Disk (music)
CD-ROM (data)
CD-I (data)
Interactive Videodisc

Write Once Optical Disks

Introduced: 1983

Function: Read and Write

Market: Precursor to Re-writable Marketplace
Other ??

Applications: ??

Standards: No

Applications for WORM. Now that rewritables are here, there are very, very few. People constantly misunderstand the fact that a WORM is for "archival purposes." A WORM just doesn't last as much as an erasable disk. And you can give me a WORM disk today, and I can alter the data for you--really simply--and you will never know.

(Laughter)

DR. FREESE: So, in terms of the market and in terms of the applications, a very, very small market--very, very, very small. And many companies have found that out, like STC, and got out of that business.

(Change of slide)

DR. FREESE: Erasables tend to be a little bit new, introduced officially in 1988. Of course, with an erasable optical disk, you can do all the things that you normally associate with magnetic disks, floppy disks, hard drives, Winchester.

There is an extremely wide variety of markets, and the applications we will go through in just a second. But you can use them just like magnetic disks. A lot of people say they are erasable or rewritable compact disks; and every once in a while, somebody will use the term: It's just like a random access tape.

(Change of slide)

DR. FREESE: Optical disks all work the same. You start out with some sort of laser source--a laser diode or whatever--take that light, put it into an optical head, and focus it onto a rotating disk memory.

You use that same laser beam to go back and read the information off the disk.

(Change of slide)

DR. FREESE: Well, that was simple enough. Why all the fuss? Why all the fuss is because it combines--uniquely combines--these particular attributes. You get a large storage capacity, which we will talk about in a minute. The disks are removable. You have increased reliability; every time you talk about more data and more capacity, reliability gets more and more and more important.

I am reminded of the time that I lost my business plan on a floppy disk, and I can't tell you why; but anyhow, I just lost it. The disk got corrupted somehow. And I was really mad and angry for about a day because it took me about a day to redo the business plan from my memory back onto the computer.

Suppose that was a little optical disk; looks just like a floppy disk, about that big [demonstrating]--only that it is going to hold 1,000 floppies. So, now, if you lose the data on that disk, that's not one day that you are going to be upset; it's 1,000 days.

So, we'll talk a lot about reliability associated with optical recording systems.

You get a random access feature associated with the disk; we will talk about the archivability and key features associated with archivability--one of the main reasons why NASA is using these disks today for some of their archival storage.

We'll talk a little bit about erasability, and I want to concentrate also a great deal on international standards.

(Change of slide)

Erasable Optical Disks

Introduced: 1988

Function: Read, Write and Erase

Markets: Medical
Imaging
Engineering
Security
Accounting/Banking
more...

Applications: "Just Like Magnetic Disks"
"Erasable CD's"
"Random Access Tape"

Standards: Yes

How Does It Work?

- * Magneto Optic Technology
- * Erasability of Magnetic Recording with the High Density, Reliability, Removability of Optical Recording

Why All the Fuss?

Combines:

- * Large Storage Capacity**
- * Removability**
- * Increased Reliability**
- * Random Access**
- * Archivability**
- * Erasability**
- * International Standard Media**

Large Storage Capacity

- * 650 to 1,000 MBytes per 5.25" Disk**
- * 16,000 to 93,000 MBytes per Jukebox System**
- * "Gigafloppy"**
- * Mainframe Storage on Desktop -- Removable**

DR. FREESE: In terms of the large storage capacity, typically you are looking at 650 megabytes on a 5.25 inch disk; that is user available storage capacity. Within the last year or year and a half, there are robotic systems, near-line storage systems; and they will provide you up to about 100 gigabytes per storage system by as many systems as you want.

A lot of people who use these disks refer to them as a "gigafloppy." The people at NASA Goddard, for example, use the disks to store the VOYAGER I and VOYAGER II data. Imagine sitting at a work station with VOYAGER I and VOYAGER II data--all the data, I understand--on just a few disks on the desk-top.

Any time you want to access any of that data, just pull down the disk, pop it in, and away you go.

(Change of slide)

DR. FREESE: The disks are removable. This is the part everybody always forgets. In fact, I can't tell you how many times a customer has come back and said: By the way, did you know these things are removable?

(Laughter)

DR. FREESE: It's a disk; you use it just like a floppy disk in that respect. They'll say: Sure, yes, I sold you the system. I knew they were removable, and it says so right there in the literature.

But people aren't used to removable data storage; so, think of it. You are sitting there are your work station; you need disk access, disk storage. So, you fill up your optical disk, or you fill up your hard drive. What are you going to do next?

Well, you can go out, and you can buy another hard drive. That is sort of an expensive solution. Or you can pull the optical disk out and just pop a new one in and set the other one aside.

(Change of slide)

DR. FREESE: So, at a work station level, what you find is virtually "infinite" storage sitting at your fingertips. Write a system once; just feed cartridges in and keep them right in front of you.

Beyond that, people discover once again the disks are removable. If you can keep them on your shelf, you can keep them on your title shelf. You can keep them in your salt mine vault in the mountains in the western region of the U.S. if you want.

(Laughter)

DR. FREESE: You can even take your data and say that university up in Michigan or a university in Ohio wants to take a look at that data, do a disk copy. Federal Express it; send it to them. If the guy happens to be on your local area network, then you can send it across the network if you wish.

My experience is that most of the time they are not linked to you. So, just do a disk copy and mail it overnight to them.

You have got the message in terms of the use; and by the way, if you stop by tomorrow, you can see one of these devices and play with it all you wish. You have, in essence, a removable hard drive.

(Change of slide)

Removability

- * Portable Cartridge
- * On-line Storage
- * Off-line Storage
- * Back-up
- * Archival
- * "Removable Hard Drive"

Increased Reliability

- * Non-contact "soft" Laser Beam
- * No Head Crashes
- * No Wear
- * No Tribology Issues

Eliminates #1 Failure for Conventional Media

Increased Reliability

- * Active Layer Buried Within Glass Substrate
- * Dust and Dirt Out of Focus
- * Reduced Susceptibility to Contaminants in Environment
- * Protective Cartridge

DR. FREESE: It is random access. I spent an hour on the phone yesterday with a guy saying: Exactly what does this thing look like? He was a PC user. And I said: Gee, it looks just like a Drive A. You plug it in. My secretary can take a system out of a box, not even reading the instructions, install the thing, get it up and working in 20 minutes. The worst part about it was that she had to teach herself which end of the cartridge goes in the slot and where the button is to push it out, to get the cartridge back out again.

It's a disk; it's a regular disk. You use it just like a regular disk. There is nothing for the user to learn or to forget. There are no new commands for you to learn; there are no new commands that you need to forget. Can you do disk copy on it? Sure.

DR. FREESE: It is rewritable. I find that people in the marketplace use the term erasability; but the technical term is rewritable, which is to say that you can delete a file and recover that space.

You can delete a file on the WORM systems, but you don't get the space back and the file will go on. And if you give me about two minutes, I can delete a file so that you never even knew it was there; there is no traceable record associated with the WORM system. All that is done on the software side.

And so, if you wish, you can put together a so-called rewritable system such that you can only record it once, and you can't ever record that spot again. So, you can have all the functionality of WORMs without any of those disadvantages.

People constantly ask me: How many times can I erase this thing? Or: How many times can I rewrite this thing? We actually know of no limit. The highest number that I know of that has ever been tested on a single track is 30 million times of rewrite/erase cycles.

You typically don't get that many cycles with tapes, floppies, or hard drives today; but the limitation is not the recording mechanism, it is associated with the tribology. You get a head crash, or you get some friction. You get some dust on a disk.

DR. FREESE: Everybody knows that optical disks come in sandwiches, like a peanut butter and jelly sandwich; you use the peanut butter and jelly layer as the recording layer. So, the actual physical recording layer is buried; it is buried underneath either glass or it is buried under bullet-proof glass. That makes it pretty tough. It doesn't make it totally perfect.

You can give one of these disks to your dog, and your dog can chew it up; and you'll lose your information. However, what this means is that the information surface is buried and that things in the environment which play a role--dust, dirt, fingerprints, sea water, whatever--stay off the information surface.

The information surface remains buried and protected; and this has great implications relative to our data storage. Put a disk on the shelf for ten years and you still want to access that data--great. In this particular case, you might even have to blow off some dirt if you want; but your data will still be there.

DR. FREESE: If you have got all this data and you have spent a lot of money obtaining that data in the first place, you want it to last for a while. You don't want it to last just a couple minutes or even a couple years. You want it to last for a very, very, very long period of time.

So, let's talk about the archivability or the stability of these types of disks and where you may end up using these instead of some other type of technology.

The first thing is what we just talked about. I don't know what a soft laser beam is, to tell you the truth; but you have a noncontact method of recording. You are not rubbing any two things together; you are not rubbing any two surfaces together.

And so, because of that, you don't have any head crashes; and you don't have any wear. You totally get rid of the tribology issues, for which there is an entire panel--I believe either later today or tomorrow--to discuss. Those issues are gone.

But those are your number one issues associated with conventional magnetic recording systems today. Because those are gone, you have increased reliability associated with these systems.

DR. FREESE: In the magnetic-optic approach, which is pretty much the standard approach today, people always ask: How about accidental erasures? In fact, the people in Washington, D.C. are always asking me: Can you take the media on the subway?

And in fact, in 1984, I had to do some tests, where I actually carried the media on the subway for a few governmental agencies.

It is important to realize that in the magnetic-optic approach, it's true that the disk has magnetic properties; but the coercivity of that disk, or how stable those domains are--how much force it takes to accidentally erase your disk--is extremely high.

(Change of slide)

Your typical, conventional magnetic recording media today is somewhere between 300 Oersteds and maybe a little over 1,000 Oersteds. My son's magnet is stronger than that. And if my son sets his magnet on top of my video tape, the information is gone.

With erasable optical disks, you have coercivities that are in the hundreds--or I should say, in the tens of thousands of Oersteds, or tens of thousands of Gauss. Those types of magnetic fields, ladies and gentlemen, don't exist in the normal environment.

There are a few Government agencies you can go to that can produce a field that strong; but typically, you can't run into this in the environment. So, you can indeed take this disk and set a magnet on it if you want; and it won't do anything. It is very, very, very stable.

And yes, it can be sent through the mail; and yes, it can be sent through the airport security checks.

(Change of slide)

DR. FREESE: Those of you who are familiar with the removable media systems, or removable Winchesters, that were a little bit more common at the beginning of the 1980s, are familiar that you could take and remove the disk pack and set it on the shelf. And if you did so and then took that and put it back in three months later or six months later, you couldn't retrieve the information.

That issue also is addressed in the optical recording systems. You couldn't retrieve that information because you got some misalignments in the head, and the tracks no longer lined up again.

With the optical disks, the optical disks are all pregrooved--little tiny grooves, just like your record album grooves--sitting in each disk. With that groove, you can have servos which follow precisely on that groove. Now, big deal--what's the importance of this to the end user?

The importance is that these disks are removable, remember. They are removable--if you want to put them on the shelf, if you want to put them in an archival vault, if you want to potentially send them to a guy in Ohio, you can. He has a different system than you do; and now, you have to deal with system to system and media to media fluctuations.

Increased Reliability

- * **Extremely High Coercivity Media
(5,000 to 50,000 Gauss)**
- * **Most Magnets too Weak to Affect**
- * **Accidental Erasure Due to Magnetic Fields Difficult**
- * **Can be Sent Through Mail and Airport Security Checks**

Increased Reliability

- * **Each Disk Hard Sectorized with Physical Grooves
(tracks)**
- * **Tracking Servo Maintains Alignment to .1 μM**
- * **Focus Servo Maintains Alignment to .1 μM**
- * **Track to Track, Disk to Disk Variations Corrected by
Servos**
- * **"Mirror" Block Corrects DC Servo
Position 2400 Times/Minute**
- * **No Long Term Drift**
- * **Enables Media Interchangeability**

Well, with these pregrooves, you can send them the media, which isn't perfectly to spec; and his system will go back and line up on that groove and store and retrieve the information in that reliable way.

Every once in a while, somebody in the audience is quite familiar with the servo mechanisms; and he says: That's great; servos will solve your problem, but your servo may drift after a while so that your laser beam isn't dead nuts on the groove any more; it's dead nuts on the edge of the groove. And isn't that a problem, too?

Well, that problem is solved also in the preformatted structure of the disks with what is called its "mirror block." What happens is: Once a revolution, the D.C. position in focus and tracking is corrected, so you get rid of long-term drift, too.

The bottom line is: You can remove these disks, and our customers do it all the time. Pull the disk out of your drive and pop it into somebody else's drive.

(Change of slide)

DR. FREESE: People always ask me about life; and to tell you the truth, the media has not been around long enough to actually quote field experience, other than for the past, say, four or five years.

People do extrapolation tests and determine the chemical stability of the disks and predict lifetimes as a result of this. Outcomes of the erasable optical disk are very, very good. I won't sit up here and quote you the numbers of the technical people because you won't believe them.

But I would like to give you a feel for what is a little bit different in the optical disks. I think you heard in the introduction that I used to be in charge of 3M's WORM activity at one point in time; and we used to do comparisons between the WORM disks and the erasable disks. The erasable disks almost always won in terms of their longer life, and usually by about a factor of 10. They did so because the erasable disks are self-passivating. What this means is that if you took a WORM disk and drilled a hole in it and put it in a very corrosive environment, eventually that corrosion would eat up the whole disk. STC constantly saw those tests.

If you do the same thing with an erasable optical disk, you can cut a hole in it and the corrosion will grow for a couple microns and will stop; it will stop growing. It is because of this self-passivation mechanism. You are familiar with them: your car bumper.

Your car bumper is made out of chromium, and chromium is a really reactive metal. It is right up there; it is one of the most reactive metals known to man. So, how come your chrome bumper stays around for so long? It does so because as soon as that chrome bumper is formed, there is a layer of chromium oxide which is formed on the outside; it oxidizes and forms chromium oxide.

Chromium oxide is really tough stuff. It's right up there with cubic zirconium in terms of stability. And so, that chromium oxide layer protects the rest of your bumper. Now, if you get in an accident and scrape that layer off, what happens? Well, a new layer forms immediately. And that same mechanism is present in the erasable optical disks. So, you have a very, very, very high or very long archival life; commercial people just say greater than ten years. At the same time, the magnetic domains are very, very stable.

If you are familiar with magnetic domains and the stability of magnetic domains, you know that on magnetic tapes, as an example, the domains move around. Well, when the domains move around, you are losing data.

Random Access

- * **File Directory Structure Stored**
- * **"Random Access Tape"**
- * **Instant Restore Applications**

And if you have ever taken an audio tape, you can hear this. You hear the music right before it starts; that's the domains moving. That is bleeding.

The ability of a magnetic media to bleed or lose its domains or have the domains move, all else being equal, is inversely proportional to the coercivity. So, here you are talking about coercivities that are ten to a hundred times higher and more stable than your conventional magnetic disk.

(Change of slide)

DR. FREESE: But the biggest thing in our compatibility is standards; that is really the issue, ladies and gentlemen. It's not just the disk that will last because some other guy is going to get up here and say: Well, my disk lasts just as long as your disk. The issue is ten years from now, 20 years from now, 30 years from now, when you pull that disk off the shelf or out of your vault, or you are wanting to call back up the old APOLLO data, or my son is going to call up the old APOLLO data, the number one issue is going to be: (1) Is that disk still intact? Yes, it will be intact. (2) Is the information still on the disk? Yes, it's still on the disk. But hey, are the players going to be around?

Are any of the companies that sell you the systems today going to be around then? That's your number one issue. Now, if you are looking to store your data for a long period of time, what is your best bet against this problem? Nobody can promise that they are going to be here 30 years from now.

Your best bet is to build standards. So, if that company is not here, at least some other company is. You see this today in all the 1 inch tapes or the 1/2 inch tapes or the 3/4 inch tapes, you know, the 6250 BPI tapes and the 1600 BPI tapes.

Why do people still use these tapes for their standards? People sell equipment, and they maintain equipment, which enables you to play those tapes.

DR. FREESE: In the optical disk world, it took us eight years to get a standard. That's a long time, believe me. Anyhow, the standard is done. There is an international standard; it has been embraced by all four international standardization committees, and it is in the 5.25 inch form factor.

There are no standards for form factors larger than 5.25. So, if you are going to store your data for archival purposes and you are looking for standards, the 5.25 inch form factor standard is done; that has been done for two and a half years now, endorsed by all of the international bodies--ISO, ECMA, ANSI, and even the classic Japan Study Committee No. 23.

This not only gives you multiple sources of commodity media, but it addresses the main issue: standards and open systems architecture, which will potentially enable you to retrieve your data. Store it on the disks now; retrieve it ten years from now. Retrieve it 15 years from now; retrieve it 20 years from now.

DR. FREESE: There are disadvantages to raise about optical recording systems; they are not all things to all people. And I'm the last guy who is going to get up here and say these systems are going to wipe out tapes or are going to wipe out hard drives or wipe out DRAMs.

All these technologies are going to coexist. They each have their advantages; they each have their strengths. And the key will just be putting them together in a hierarchy. Disadvantages associated with this industry. The technology typically requires fairly high manufacturing tolerances. So, the cost for the home user, the clerical station, is too much; it's too high.

(Change of slide)

Archivability

- * Self Passivation Mechanism
- * Expected Life > 10 Years
- * High Coercivity Stabilizes Magnetic Domains
- * Short Term or Long Term Memory

#1 Failure: Cartridge Abuse

International Standard Media

- * 5.25" Form Factor
- * ISO, ECMA, ANSI, JC23 Ratified
- * Multiple Sources of Commodity Media
- * Random Access to Archival Data
- * Mainstream User Acceptance

Erasability

"Just Like Magnetic Disks"

"How Many Erase Cycles?"

<u>SYSTEM</u>	<u># CYCLES</u>	<u>LIMITATION</u>
Tape	1,000's	Friction, Wear
Floppy	100,000's	Friction, Wear
Hard Disk	100,000's	Tribology, Head Crash
Optical	10,000,000's	???

Disadvantages

- * Technology Requires High Manufacturing Tolerances**
- * Cost too High for Home User, Clerical Station**
- * New Technology, Long Implementation Cycle**
- * Not a Winchester Replacement
(not fast enough for some applications)**

The cost relative to a hard drive is not a valid comparison. You will hear speakers constantly sit there and say: Let's compare a hard drive with an optical drive. Okay. They are almost two different animals.

A hard drive comes with a fixed capacity; it's a high performance device. An optical drive comes with virtually infinite capacity; you just keep on feeding those cartridges in there. It's a removable device. They are both disks.

One is not going to replace the other; and the other is not going to replace the one. Those things are going to coexist.

But it is a new technology; the industry is just starting. And I would just like to remind everybody that the optical disks are not necessarily a Winchester replacement or a hard drive replacement, nor are they necessarily a tape drive replacement.

But they can combine the best features of all these creating "a random access tape" or a removable hard drive.

(Change of slide)

DR. FREESE: In terms of where the industry is today, I'll start off with the smallest size. 3.5 inch subsystems were just announced, and the media standards are in essence completed. This is a this year type of activity. 3.5 inch is going to hold 127 megabytes on a disk; it will be about \$2,000 to \$3,000 a subsystem. And really, this is aimed at the PC world.

Having addressed NASA audiences before, 127 megabytes a disk, I don't think, is quite enough capacity for most people sitting in the audience. But the 3.5 inch disks have just been introduced; and so, you have seen many announcements, I think, within the last three months associated with these.

(Change of slide)

DR. FREESE: In terms of system solutions, 5.25 inch and robotic systems, or so-called jukebox systems, offer really high capacity systems, and complete system solutions. You the user don't have to know anything about how this device works; but you simply use it on a PC just like a big Drive A. You use it on a DEC just like another DU device. You use it on a Sun microsystem just like another SD device.

It is totally transparent, no new commands to learn or old ones to forget.

You have solutions for almost all your popular computers nowadays, an entire line of DEC computers, the DEC network, all your Sun computers and Sun Sparcs.

DR. FREESE: The old Suns, the Sun NFS network, the new IBM RS6000s, all your PCs and their clones, the Novell networks; in the graphics industry, you often run into Scitex whisper stations, Macintosh, Appletalk.

(Change of slide)

So, you have got complete storage systems for almost all of your mainstream computer systems.

(Change of slide)

DR. FREESE: State of the industry for 5.25. Those standards are complete and being implemented worldwide. Commercial products began about two to two and a half years ago, where we talked about 650 megabytes as the increment on a cartridge. Put as many cartridges in your system as you want.

State of the Industry: 3.5" Subsystems

- * **Media Standards Complete**
- * **Commercial Products Begin 91-92**
- * **127 MByte/Disk**
- * **5-10 MBits/sec**
- * **30-50 msec Average Access**
- * **\$2,000 - \$3,000/Subsystem**
- * **P.C.'s Only**

State of the Industry: 5.25" Subsystems

- * **International Media Standards Completed**
- * **Commercial Product Shipments Began 8/88**
- * **650 MBytes User Available/Cartridge**
- * **5-10 MBits/sec**
- * **30-100 msec Average Access**
- * **\$4,000 - \$7,000/Subsystem**
- * **First Applications Workstation Oriented**

Present Status: 5.25" System Solutions

- * COMPLETE SYSTEM SOLUTIONS EXIST FOR
MOST COMMON COMPUTERS
E.G., ALPHATRONIX INSPIRE SERIES AVAILABLE
FOR :**

DEC VAX COMPUTERS

DEC MICROVAX COMPUTERS

DEC VAXSTATION WORKSTATIONS

DEC DECSTATION (RISC BASED)

WORKSTATIONS

DECNET NETWORKS

SUN SPARCSTATIONS, INCLUDING SPARC 1

AND SPARC 2

Present Status: 5.25" System Solutions (cont)

SUN 2,3,4 WORKSTATIONS

SUN NFS NETWORKS

IBM RS6000 WORKSTATIONS

IBM PC-AT AND PC-AT CLONES

IBM PS-2

NOVELL NETWORKS

SCITEX WHISPERSTATIONS

APPLE MACINTOSH COMPUTERS

APPLETALK NETWORKS

People who use Sun computers quite often sort of discover the removability aspect. Sun people, in particular, are constantly running out of disk space. So, when you run out of disk space, pop it out and put a new cartridge in; and you have another 650 megabytes.

Looking at about 5 to 10 megabits a second on our optical drive, 30 to 100 milliseconds of average access time; it costs about \$4,000 to \$7,000. And an awful lot of the 5.25 inch are really work station oriented. You see this in your Earth Resources in your NASA applications, where NASA may receive the information from some satellite orbiting; but there are an awful lot of people who do analysis on that information, universities in particular. So, NASA may pull the information down on DEC systems and then want to distribute these systems to their PC customers, if you will.

(Change of slide)

DR. FREESE: So-called jukebox systems or auto changers, robotic loaders. You can take these 5.25 inch disks and stick them in a rack essentially and have an auto changer, albeit a very, very fast one; and now you just simply expand by the number of cartridges the amount of on-line data you can get.

So, typically in a little box about yea big [demonstrating], you are looking at about 100 gigabytes or so of on-line disk random access storage.

Commercial products began shipping about a year and a half ago. You can get as small as you want I guess, but anyhow, almost up to 100 gigabytes, user available on the system.

The speed of the robot typically is quoted in a few seconds.

Today, there are mostly vendor unique solutions in the jukebox arena. Be very, very careful about this because if your application is archivally oriented, do you want to store your data on an ISO standard disk using nonstandard disk directory structures?

Do you want to bet your database on one company? Most people say no. If the answer is no, then make sure you have an open systems architecture solution. Make sure that that disk is always standard, not proprietary.

If you are on a DEC system, make sure your disk is always a standard DEC disk; it is not a vendor unique disk.

Make sure that you can take that disk and send it anywhere in the world you want, and anybody can read and write to it, following the standards. Again, you have got your most popular platforms.

Where is the state of the industry in optical tape? I didn't want to preempt the talk by Bob McLean tomorrow from ICI; but anyhow, let me review very quickly optical tape.

(Change of slide)

DR. FREESE: The advantages of optical tape are mainly optically derived; that is, there are very, very high densities that you get associated with optical recording and the fact that you are using a laser beam to rewrite and erase.

You can make tapes basically as long as you want. Optical tapes typically talk about gigabytes and sometimes even terabytes on a given reel of information.

(Change of slide)

State of the Industry: Jukebox Systems

- * **Commercial Product Shipment 89**
- * **6 to 93 GByte/System**
- * **Disk Exchange - 2-3 Seconds**
- * **Mostly Vendor Unique Solutions**
- * **Some Open System Architecture Solutions
e.g. "Alphatronix Inspire"**
- * **Dec, Sun, PC, Macintosh, IBM Platforms**
- * **Decnet, N.F.S., Novell, Apple Talk Networks**

Optical Tape

- * **Advantages: Mainly Optically Derived**
 - **Same High Density as Optical**
 - **Higher Capacity: Gigabytes to Terabytes**

Optical Tape

- * **Disadvantages: Mainly Tape Based**
 - **Tape Stretch/Break**
 - **Tribology/Friction**
 - **Contact Recording**
 - **Environmental Stability**
 - **Non-Random, Sequential Access**

DR. FREESE: In a nutshell, the disadvantages are associated with the fact that it's a tape. Tapes are sequential access; tapes break; tapes are flexible. It is a contact method of recording because now you have got to spool your tapes. You have surfaces that are rubbing now. So, you have all the issues associated with tapes that need to be addressed.

(Change of slide)

DR. FREESE: In terms of applications, I would like to talk about two in particular; but having talked with NASA Goddard for a couple of years now, this is one of your most common issues that I have been able to discern.

And that is: Here you have this disk; it stores 650 megabytes. Great. You can put your VOYAGER I data on it; you can put your MAGELLAN data on it; you can put your shuttle data on it. You can put your images on it.

But you have so many people who want to analyze that data; and NASA has so many relationships with universities and private corporations and is constantly shuttling this data around. They have a really big problem, and that is that I may be NASA and I store my information on a DEC system; but the guy who is going to analyze that data is sitting on a Sun system. Well, the disks are portable; so, I can take my DEC disk and give it to my Sun guy. And say: Here, here's the data; go ahead and have at it. Except that's a Sun disk; he's on a Sun system, and that's a DEC disk. Those two things are incompatible.

Another example of that is: I'm NASA and I sit down at a DEC system, sending my data out to universities. The people at the universities are using PCs to analyze their data. Can you take a DEC disk and put it on a PC? The answer to that question obviously is: No, you can't.

See, now, the optical disk enables at least that possibility because you can store the information on that disk; it is high capacity, and you can ship it around because it's removable.

(Change of slide)

DR. FREESE: Bypass is a solution to that problem. Bypass is an application software program which has been designed explicitly for this problem. And what it enables you to do, just like the picture implies, is that you can sit on a PC machine using DOS and access standard DEC files or standard DEC disks.

And so, for the first time, you've got the possibility of a universal storage medium. You in NASA go ahead and store your information on either DEC, Sun, or PC systems; and that information can be distributed to anybody else who might be on a DEC, Sun, or PC system, at all times keeping the disk in standard file structure format.

So, you don't get any disadvantages.

(Change of slide)

DR. FREESE: So, with Bypass, you've got the potential application of a universal storage medium, adhering to the international standards, adhering to the file structure standards at all times. At no time do you have a unique disk--a unique NASA disk, a unique Alphatronix disk, a unique anybody's disk.

The cartridges are easily interchanged; just pull them out from one and stick them in the next. And there is nothing for the user to do; he just simply has access to those files.

DR. FREESE: Bypass then is available for VHF, Sun and DOS as of today--oops, I see we are going to do MAC in the future.

Optical Tape

Status:

See Wednesday Talk

Sample Earth Resources Applications

- * **Voyager I, II data analysis (NASA Goddard)**
- * **Magellan Image Analysis (NASA Goddard)**
- * **Space Shuttle Mission Photography
(NASA Huntsville)(Hughes)**
- * **Weather Satellite Photos/Analysis
(NOAA)(Fleet Numerical)**
- * **Oceanographic Mapping (Woods Hole, Oregon State)**
- * **Data Distribution
(Marshall SFS, USSD, JPL, TRW, Aerospace)**

Rewritable Optical Disk: The Universal Storage Medium

- * International Standards Adopted and Implemented**
- * Cartridges Easy to Physically Interchange Between Hosts**
- * Bypass Platform Dependent File Systems**
- * Requires No Modification to Access Data on any Host**

Data Interchange Using Optical

A Solution Presents Itself:

- * Rewritable Optical Technology**
- * 650 Mbytes per Cartridge**
- * Easily Portable**

Relative to an archival application, this is also sort of interesting because there is a constant migration of computer systems, a trend that I think everybody in the room can identify. What a person used to do on a mainframe, he now does on a work station. What he used to do on a work station, he is now doing on a PC. So, tell me: Where do you store your information? How do you store your information? Do you store it in DEC format? Do you store it in Sun format? Do you store it in PC format? Which one do you think you ought to do?

With Bypass, it doesn't matter. Store it in one of the three, and you will be able to access one of the three at any time in the future.

DR. FREESE: Other sample applications. Some of the systems are Goddard; I'm aware of at least two of them. One is people who are doing the analysis on the VOYAGER I and VOYAGER II data; and actually, I was surprised, but very pleased, to find out that VOYAGER I and VOYAGER II actually are still sending us data. And it is under constant analysis.

The MAGELLAN program also uses various optical and storage devices. The space shuttle dumps its imagery to jukeboxes down in Huntsville.

(Change of slide)

NOAA has been using the optical storage systems for their weather satellite and weather image analysis, as does the Navy for their weather forecasting.

Woods Hole and Oregon State are a couple of examples where they are doing more photographic mapping, although in this particular case it is of the ocean. And the Marshall Space Flight Center, along with a few other people, are now really getting interested in the data distribution application, where you can take the disk--be it on DEC, Sun, or PC format--and distribute it across the country or across the world to anybody else on one of those three particular platforms.

And then, they can do their own analysis on the data and communicate that to the rest of the world.

These are a few of the classic applications associated with earth resources.

(Change of slide)

DR. FREESE: Other applications include anything to do with images whatsoever--scanned documents or images, CAD/CAM documents or images. If you have ever had a CAT scan or an MRI scan, they are used to store those particular images.

DR. FREESE: We already talked about satellites and geologic data analysis.

(Change of slide)

DR. FREESE: Data logging and analysis is sort of interesting. Here is an application where people used to use tapes, but they found that the tapes weren't archival. You couldn't keep your tapes around for a long time; the tapes had binders on them, chemicals and glues. The tapes stick after a while. You see that in your NASA tapes from the 1960s where, when you unroll the tape, particles fall apart; and the thing pulls apart.

Now, the pharmaceutical industry is converting over to erasable optical disks just to solve this problem mandated now by the FDA, where they have got to store all their data; and they must keep it on-line, and they must keep it for seven years. And they have to keep their original data.

That's an application where they decided, gee, archival life is the number one issue here; and so, therefore, we are going to store these things on standard optical disks.

The Bypass™ Solution

- * Developed for Rewritable Optical Disks**
- * Works with Many Technologies**
- * Copies Information To/From Foreign Disk Types and Formats**
- * Uses Simple Commands**
- * Reads/Writes Cartridges From Any Supported System**
- * Easy to Add Other Hosts Quickly**

Advantages of the Bypass Solution

- * Allows Access to Data in Native File System Format**
- * All File System Commands Work Normally**
- * Requires No Modification to Native File Systems or Applications**
- * No Need to Intercept Requests and Translate**

Bypass

- * **Device Independent**
 - VMS**
 - SunOS**
 - DOS**
 - MacOS (future)**
- * **Converts DOS 3.x to 4.0**

Bypass Applications

- * **Share Data Between Field and Corporate Offices**
- * **Data Distribution**
- * **Move and Release Source Code and Programs**
- * **Distribute Disks with Different Formats**

Archiving Applications

- * **Key Issues:**

- **Ability to Retrieve & Store Data in Future**
- **Systems Must Be Supported/Serviced**
- **Media Life > 10 Years**

Archiving Applications: The Inspire Solution

- * **Open System Architecture**
- * **Adheres to All ISO/ANSI International Standards**
- * **Adheres to all Native File Format Standards**
- * **Media Life > 10 Years**

Sample Applications

- * **On-line Data Storage**
- * **Back-up Data Storage**
- * **Archival Data Storage**

Sample Applications

- * **CAD/CAM**
- * **Image Processing**
- * **Medical Imaging**
- * **Scanned Document Storage**

Sample Applications

- * **Computer Integrated Manufacturing (CIM)**
- * **Geologic Data Analysis**
- * **Satellite Image Analysis**

Sample Applications

- * **Software Distribution**
- * **Data Distribution**
- * **Data Logging/Analysis**

Sample Applications

- * **Government/Security**
- * **Desktop Publishing**
- * **Software Development**

Future Directions - Now to '94

- * **3.5" Media Standard Completed
Products Announced
~ 150 MBytes/Cartridge**
- * **5.25" Disk: Two to Four x Present Capacity
(1-4 GBytes/Disk)(Backward Compatable)(Standardized)**
- * **2 to 5x Increase in Transfer Rates
(20 to 50 Mbits/sec)**
- * **15-25 msec Average Access**
- * **Half-Height Drives**

I should mention the issue of erasability since it was brought up by the prior speaker; but that's what this industry uses today. They use tapes; tapes are rewritable. Tapes are erasable; tapes are correctable.

Only microfilm and microfiche is considered not to be correctable.

So, for one to switch from tapes to optical disk was no issue whatsoever.

DR. FREESE: There are a tremendous number of systems in Washington, D.C. all associated with Government applications and all associated with the removability aspect of these cartridges. You can keep them on-line in your computer if you want; but when you are done with that data, you can pull it out and stick it on the shelf. You can keep your system right there. If you wanted to send it to somebody, you could.

DR. FREESE: I would like to take the last few minutes and talk a little about future directions. I think it's important to realize that all technologies are advancing.

I remember at the beginning of the 1980s hearing somebody stand up at an optical disk conference and say: This stuff is just going to wipe out the hard drive industry. And I remember thinking: Well, I'm not too sure about that. I've seen just the opposite.

I've seen magnetic disk people stand up and say: You know, tape is going to wipe out the optical disk industry, or whatever. They are all different solutions. They all have different advantages. And it's a mistake to assume that the technologies aren't advancing; they are.

Five years ago, a lot of people rang the death knell for tapes, saying tapes just can't go much further. Yet at the same time, in our own laboratories, we were storing tapes in a helical scan format with densities a factor of 10, a factor of 20--higher than what anybody was talking about then.

So, we have to be very, very careful about industries that aren't advancing. I think it's a good assumption to assume that they all will advance.

What I would like to talk about here is advancements that you will see, I think, in the optical recording industry between now and the immediate future in a commercial sense.

I won't talk technology because technology isn't a solution to your problem.

Between now and 1994--in fact, the very first thing that is happening this year is your 3.5 inch formats are available; and in the coming years, what you will see is all your PCs that are coming out of Japan will have a 3.5 inch optical drive or optical drive option associated with them.

On the 5.25 inch, the issue isn't technology. I can think back to 1983 when, in the laboratories, we took a standard 650 megabyte optical disk and stored 4 gigabytes on that disk, just using the same system that we had back then. The issues are not technology driven in the optical world; they are standards driven.

It took us eight years to get a standard in erasable optical recording. It then took another four years to get a standard in the 3.5 inch form factor; and now, the standards committees are turning their attention right back again to the 5.25 inch.

The 5.25 inch will evolve then, and discussions right now are anywhere between two and four times the present capacity per disk. So, somewhere between 1 and 4 gigabytes per disk on the second generation erasable optical 5.25.

The key part of the standards is that it be backward compatible with your existing disks. And the good news, at least from our perspective and probably from this audience's perspective, is that everybody that we know of talking in the standards bodies is making the standard backward compatible.

So, whatever disks you have got today will be able to be played on the second generation systems coming up.

When this occurs is a standardization issue; it is not a technology issue. The technology to do this is getting to be about seven years old. That's good because it makes it good, hard, and stable. But this is an issue of standards.

In my own personal opinion, that standardization will take about two more years because it tends to be a little bit political.

Performance will increase in terms of transfer rates, average access times, and getting into 1/2 height types of form factors.

(Change of slide)

DR. FREESE: There is a lot of talk about blue/ green laser diodes and taking the existing technology and changing the head, changing the colors of the head, and getting a 4X increase in capacity. And all that's basically true.

The key issue relative to the green laser diodes, which will then give you somewhere around 6 gigabytes on a 5.25 inch disk removable is commerciality. There are such diodes, I understand; in fact, I've understood that they have been around for quite a while, but they are not reliable yet. They are not commercially available yet.

But between now and the year 2000, those things will become available; and when they do, you'll see the same disks--the same optical disks--take about a 4X leap in their capacity.

Some interesting work which has not progressed in the commercial world is multiple beams. All of our technologies today use a single beam to read, write, and erase. But in 1983, 1984, and 1985, we also sponsored work at RCA, where they took 18 beams and recorded all those beams in parallel simultaneously. Well, the media will do that; the technology will do that, as soon as laser diode arrays are available in that type of size. Then, you will see another quantum leap in the data rates.

Finally, between now and the year 2000, there is a great deal of work going on in lightweight heads--very, very compact heads, heads that look about the same size as a computer chip or as magnetic heads today.

So, you can expect between now and the year 2000 to see access times close the gap in the mechanical sense between today and whatever exists for hard drives. Now, hard drives will be moving; hard drives today are getting into the single digits of milliseconds, but so are these heads. There is no reason that they shouldn't.

(Change of slide)

DR. FREESE: So, summarizing, and in conclusion, the erasable optical recording industry is just starting. It is a new industry. We have found it so far to be a brand new tool for the work station user. It doesn't eliminate or supplant tapes; it doesn't eliminate or supplant hard drives or floppy disk drives.

What it is, is an additional option in the hierarchy of storage solutions for the user. There is a tremendous growth path for this particular technology; and usually, the main

Future Directions - Before Year 2000

- * Green Laser Diodes**
Capacity up to 5-6 GBytes/5.25" Disk
- * Multiple Beam Laser Diodes**
Data Rates 80-160 Mbits/sec
- * Lightweight Heads (Holographic, Luneberg)**
Access Times Similar to Hard Drive

Summary

- * Erasable Optical Industry Just Starting**
- * New Tool for Workstation; Networked User**
- * Significant Growth Path**
- * Wide Variety of Applications, Many New**

reason why companies have gotten involved is for this particular growth path. And a lot of the applications which we are seeing for optical storage today are really just starting.

People are rediscovering the joys of removability, the joys of standardization and open systems architecture, and the joys of random access tape devices.

So, thank you very much for being a real patient audience. If you have any questions, do we have a few minutes for questions?

MAGNETIC DISK

John C. Mallinson
MALLINSON MAGNETICS, INC.
7618 Reposada Drive
Carlsbad, CA 92009

Magnetic disk recording was invented in 1953 and has undergone intensive development ever since. As a result of this 38 years of development, the cost per byte and the areal density have halved and doubled respectively every 2-2 1/2 years. Today, the cost per byte is lower than 10^{-6} dollars per byte and areal densities exceed 100×10^6 bits per square inch.

In this talk, the recent achievements in magnetic disk recording will first be surveyed briefly. Then the principal areas of current technical development will be outlined. Finally, some comments will be made about the future of magnetic disk recording.

PRESENT ACHIEVEMENTS

High end disk drives today operate at areal densities of between 50 and 100×10^6 bits per square inch, with, typically, 2500 tracks per inch and 30,000 flux reversals per inch. When "run length limited" coding is used, the effective linear bit density is 40,000 bits per inch. Areal densities tend to be higher the smaller the diameter of the disk.

Data rates run as high as 6 Megabytes per second (48 Megabits per second) per single head-disk channel. Parallel access disk systems, with as many as 10 heads in parallel have been manufactured which provide the full CTIR 4:2:2 component digital video output rate (216 Megabits per second).

Since as many as 6 disks can be fitted in the standard 5 1/4" full height form factor package, 5 1/4" drive data capacities exceeding 2 Gigabytes are now available from several manufacturers.

In summary, it may be said that the magnetic disk products being manufactured today offer access times, data rates and drive bit capacities considerably in excess of those offered by optical disk drives. Areal density is the only parameter which currently falls below that of optical disks, by a factor of 3-4.

AREAS OF TECHNICAL DEVELOPMENT

The overwhelming success of magnetic disk products over the last three or four decades has led to the establishment of a \$50 billion per year world wide business in disk drives. This enormous business supports research and development into every conceivable aspect of disk recording technology in order to permit continuing increases in performance. Only the major areas of such research and development can be discussed below.

IMPROVED RECORDING MEDIA

Virtually all modern disk drives now use thin film metallic media with coercivities close to 1000 Oe. It may be expected that coercivities exceeding 2000 Oe will be used in the next few years. Higher coercivities lead to both sharper output pulses of greater amplitude and also to improved signal-to-noise ratios.

IMPROVED WRITING HEADS

As the medium coercivity increases, it is necessary to increase the saturation induction of the writing head pole tip materials. Presently, Alferil and Permalloy with maximum

inductions of 10-12,000 G are used. Materials such as Co-Ru and Fe-N with maximum induction of 16,000 and 19,000 G may be expected to be introduced.

NARROWER TRACKWIDTHS

It has been realized for two decades that, when seeking higher areal densities, it is better to use narrower trackwidths than higher linear densities. Operation with trackwidths substantially narrower than normal (10 μ m) leads to a number of very fundamental questions concerning the operation of the track following servo system. In particular, the outstanding question is "what is the source of the tracking error signal?". In magnetic disks today the source is a previously written magnetic disk servo track and it is only possible to operate the tracking servo when reading but not when writing. In optical disks, which operate at 5-6 times the track density, the source is always some physical feature (pits, grooves, bumps, etc.) and the tracking servo system is then operable during both reading *and writing*. This leads to another question: "Will magnetic disks eventually use optical tracking servo systems?".

IMPROVED READING HEADS

As trackwidths decrease, it becomes increasingly difficult to keep the channel signal-to-noise ratio media-noise limited because the output voltage of an inductive head falls proportionally with the trackwidth. It is anticipated that the next generation of high end disk drives will use magneto-resistive (M-R) reading heads where the magnetic fields from the medium changes the electrical resistance of a thin film M-R element. Considerably higher output voltages are available with M-R heads and they are independent of head-medium relative velocity.

IN-CONTACT OPERATION

Today's disk drives operate with a deliberate head-to-disk spacing of, typically, 6-8 microinches (0.15 μ m). It is known that both the writing and reading processes on magnetic disks improve when the spacing is reduced. All disks today are overcoated (Ag-Sn, ArO₂ amorphous C, ZrO₂, etc.) in order to control friction and wear and it seems very likely that, together with redesigned heads of significantly lower mass, continuous operation in contact may become possible. This is particularly true at low head-to-disk relative velocities.

SMALLER DISK DIAMETERS

An interesting sequence of design changes becomes possible following a reduction in the head-to-disk spacing. First, a higher linear density may be written. Second, because the data rate has now become too high, the disk diameter or spindle RPM must be reduced. Third, at the reduced head-to-disk velocity, it now becomes possible to reduce the head-to-disk spacing even further because any mechanical impact now transfers less energy. Fourth, if a smaller disk diameter has been chosen, the mechanical tolerances (flatness, areal runout, etc.) are reduced which again permits the head-to-disk spacing to be reduced even further. This sequence has led the drive industry from 5 1/4" to 3 1/2" to 2 1/2" to 1 1/2" diameters with increasing areal density. Still smaller diameters and higher areal densities are anticipated.

As an example of the levels of performance attainable when many of these developments are combined, consider the 1989 IBM 1.1 Gigabit (1100 Megabit) per square inch technology demonstration:

- Medium coercivity - Cobalt-Platinum - 1700 Oe
- Write Head-thin film-trackwidth 4 μ m
- Read Head - magneto-resistive - trackwidth 2-3 μ m
- Head-to-disk spacing - about 1 microinch
- Linear density - about 160,000 bpi
- Track density - about 7,000 tpi

With this demonstration, IBM showed that magnetic disk recording has the potential to exceed today's optical disk areal densities by about a factor of 2.

THE FUTURE

The IBM 1989 demonstration proved 1.1 Gigabit per square inch feasibility. Today's research papers (see, for example Intermag '91 paper MA-01) discuss demonstrations of 2 Gigabit per square inch (at 17,000 tpi and 120,000 frpi). It seems to be abundantly clear that the magnetic disk recorders could range from the 50-100 Megabit per square inch of today's manufactured hardware to future products with areal densities perhaps as high as 16 times greater.

It used to be said that the great advantage of optical (versus magnetic) recording was that it was not necessary to fabricate anything with dimensions comparable to the wavelength of light in order to achieve very high areal densities because the lens could focus the light down to Lord Rayleigh's' diffraction limit.

Nowadays, it seems that a very fundamental change in philosophy has occurred. Indeed, it is frequently stated that the real advantage of magnetic versus optical recording lies in the fact that the only effective limits operating today concern just how small can certain features and objects be made and that their dimensions are not limited by mere physical diffraction of light! For example, the gap-length in mass-produced 8 mm VCR heads is 10 microns, which is but one third the wavelength of red light.

The steady increase in areal density, by a factor of 2 every 2-2 1/2 years, has been mentioned already. By this criterion alone, it appears then that magnetic disk recording technology can sustain another 20 years of growth (a factor of $16 = 2^4$; $4 \times 2.5 = 10$ years) on the basis of demonstrables which exist in the laboratories today.

To move from scientific extrapolation to the realm of technical speculation, it seems to be very likely that 1 Gigabit (10^9) per square inch areal densities will appear in disk (and video tape) drives in considerably less than 20 years. Indeed some industry observers have opined that 5 1/4" full height drives with 100 Giga-byte capacity will appear before the year 2000: this represents a doubling of the historic rate of increase. Given the magnitude of the research and development activities in magnetic disk recording being undertaken worldwide, even such surprising estimates do not appear, to this writer, to be unduly optimistic!

AREAL DENSITY

$$= \frac{\text{TRACK DENSITY}}{(\text{TPI})} \times \frac{\text{BIT DENSITY}}{(\text{BPI})}$$

VOLUME DENSITY

$$= \text{AREAL DENSITY} \times \frac{1}{\text{THICKNESS}}$$

TYPICAL AREAL DENSITIES, CAPACITIES AND COSTS TODAY

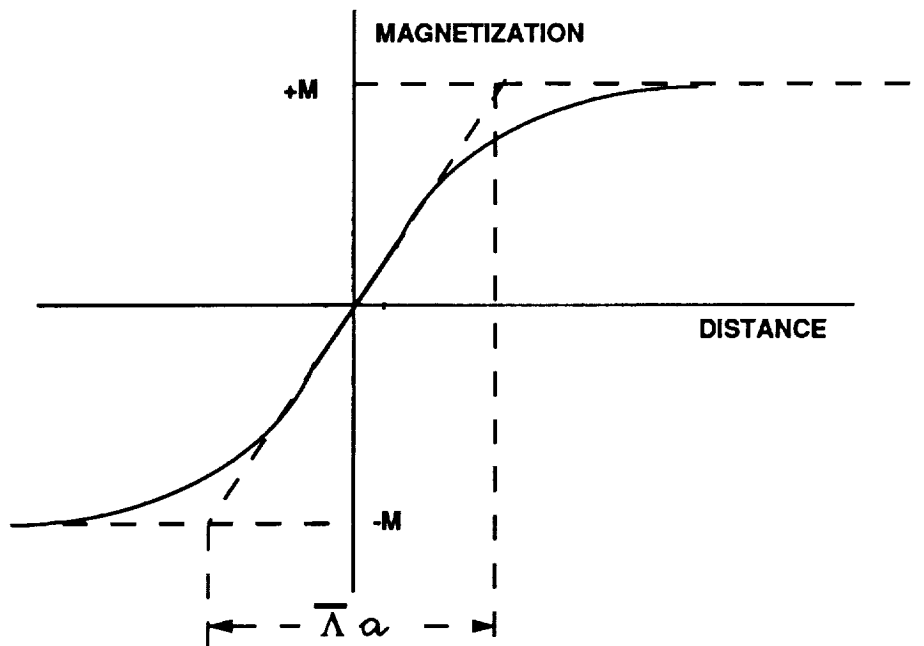
	AREAL	CAPACITY	COST
HI-END (LARGE)	60 10 ⁶	10 10 ⁹	\$100K
HI-END (SMALL)	100 10 ⁶	2 10 ⁹	\$2000
HI-END (SMALL)	20 10 ⁶	0.1 10 ⁹	\$400

NOTE: 10⁻⁴ CENTS/BYTE
10⁻⁵ CENTS/BIT

AREAL IS IN BITS/SQUARE INCH

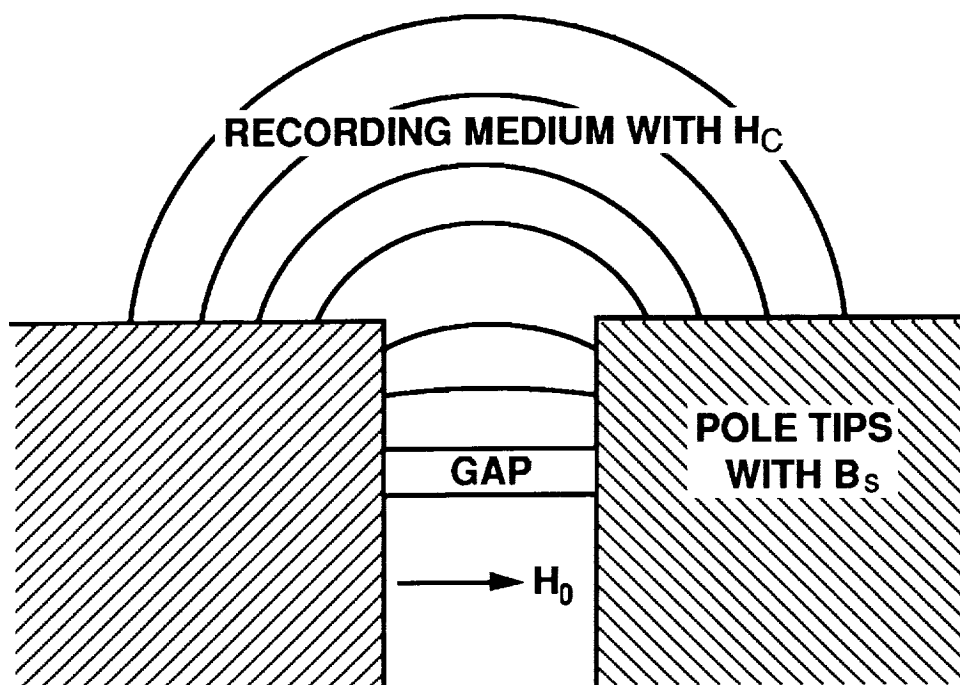
CAPACITY IS IN BYTES

EFFECT OF RAISING H_c



$$a \propto \sqrt{H_c}$$

DATE	H_c (OE)
1980	300
1986	600
1989	900
Now	1200-1400
Future	2800 Reported in Literature

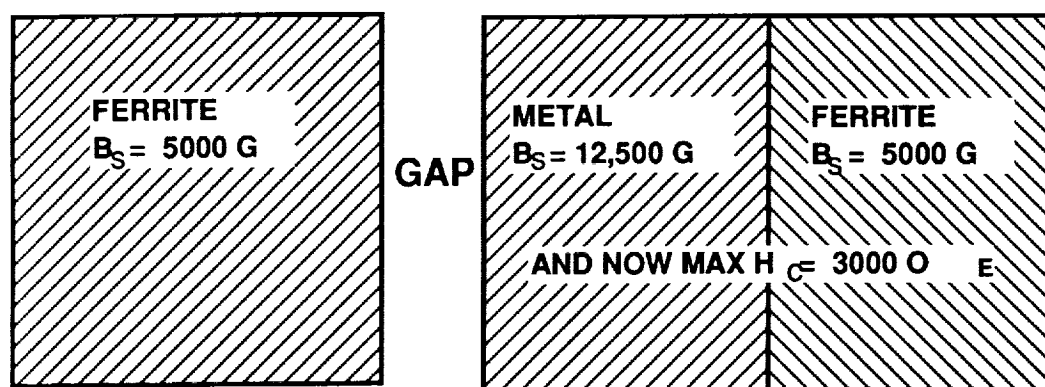


$H_0 \approx 2.5 H_c$ (TO WRITE PROPERLY)
 $H_0 \leq 0.6 B_s$ (TO AVOID POLE TIP SATURATION)

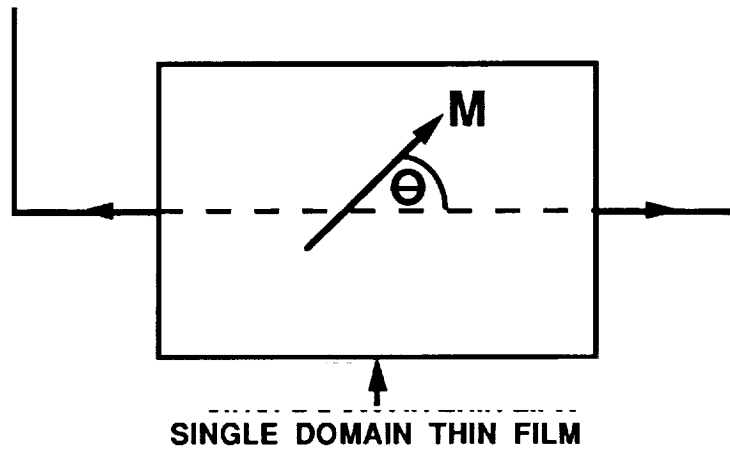
FOR EXAMPLE: WITH FERRITE POLE TIPS $B_s = 5000 \text{ G}$

$0.6 B_s = 3000 \text{ G}$
 $\therefore \text{MAX } H_c = 1200 \text{ Oe}$

METAL-IN-GAP HEADS



MAGNETO-RESISTIVE HEADS



RESISTANCE VARIES AS $(\cos \theta)^2$

$$V (\text{MRH}) = 20,000 \frac{V (\text{INDUCTIVE})}{\text{SPEED} \times \text{NO OF TURNS}}$$

SUPPOSE DISK SPEED IS 200 IPS:

$V (\text{MRH}) = 20,000 \times V (\text{INDUCT})$	FOR	$N = 1$
$= 2,000$		$N = 10$
$= 200$		$N = 100$
$= 20$		$N = 1000$

MRH ARE THE KEY TO HIGHER TPI's WITH SMALL DIAMETER DISKS.

"IN CONTACT" DISKS

THE TRACK FOLLOWING ISSUE

OPTICAL DISKS

ATTAIN 10-15,000 TPI

ALL USE DIFFRACTION OF LIGHT OFF
MECHANICAL FEATURES

MAGNETIC DISKS

CURRENTLY ATTAIN 2-3,000 TPI

ALL USE A FEATURELESS SURFACE
WITH MAGNETIC SERVO TRACKS

IDEALLY, THE SOURCE OF THE TRACK FOLLOWING SERVO SIGNALS SHOULD BE INDEPENDENT OF THE RECORDED DATA, SO THAT THE TRACK FOLLOWING SERVO CAN OPERATE AT ALL TIMES.

WILL MAGNETIC DISKS ADOPT OPTICAL TRACKING?

SMALL DISK EVOLUTION

- A) MAKE DISK SMOOTHER
- B) LOWER FLYING HEIGHT
- C) INCREASE LINEAR DENSITY
- D) REDUCE SPEED
LOWER RPM
SMALLER DISK
- E) SMALL DISK IF FLATTER
∴ LOWER FLYING HEIGHT AGAIN
- OR F) ACCELS LOWERR AT LOWER RPM
∴ LOWER FLYING HEIGHT AGAIN
- G) REPEAT CYCLE (C)

IBM'S 1 GIGABIT/INCH² DEMONSTRATION

DECEMBER, 1989

Metallic Thin Film Disk	HC = 1800 Oe δ = 300A Co-Pt-Cr
Inductive Thin Film Write Head	W = 4 μ m } 6-7000 TPI
Magnetoresistive Read Head	W = 3 μ m
50% Roll-Off Density	= 110,000 FRPI
Linear Bit Density	= 160,000 BPI
Data Rate	28 Mbs
Error Rate	= 10 ⁻⁸ - 10 ⁻⁹
Signal-to-Noise Ratio	= 23 dB
Areal Density	= 1.18 10 ⁹ Bits/Inch ²

$$\frac{\text{SIGNAL POWER}}{\text{NOISE POWER}} = \text{NO OF GRAINS/BIT CELL}$$

IBM'S DEMO HAD 500 Å (2 μ INCH) METALLIC GRAINS

THE BIT CELL WAS ABOUT 6 μ " LONG X 120 μ " WIDE

$$\therefore \text{NO OF GRAINS/BIT CELL} \approx 200$$

$$\therefore \text{SNR} = 200 \text{ OR } 10 \log_{10} 200 = 23 \text{ dB}$$

- * This demonstrates that the basic laws governing recording hold to Giga-bit densities.
- * Note that a doubling in areal density costs a halving (-3 dB) in SNR with thin media.

THE FUTURE
SUPPOSE 2X AREAL DENSITY
EVERY 2 1/2 YEARS HOLDS

<u>DATE</u>	<u>AREAL DENSITY</u>
1991	100 10^6
1994	200 10^6
1996	400 10^6
1999	800 10^6
2002	1600 10^6

This evolution appears to be feasible on the basis of today's laboratory demonstrations.

A PROFOUND CHANGE IN DESIGN PHILOSOPHY

OLD CREDO

The advantage of optical recording is that dimensions comparable to optical wavelengths do not have to be manufactured.

NEW CREDO

The advantage of magnetic recording is that the effective limits are governed by how small can objects be made and this limit far exceeds optical diffraction.

MAGNETIC TAPE

Harriss Robinson

Datatape Incorporated

ABSTRACT

The move to visualization and image processing in data systems is increasing the demand for larger and faster mass storage systems. The technology of choice is magnetic tape. This paper briefly reviews the technology past, present, and projected. A case is made for standards and the value of the standards to users.

INTRODUCTION

Major changes are occurring in data storage and processing systems technology. It is hard for most of us to keep up with what is happening in our own fields, let alone what is going on in related disciplines. Even when we can see the general direction of change, it is difficult to forecast the timing and implement systems which bring all elements together at the same time to provide optimum performance, at a reasonable cost.

One fundamental change affecting all of us is the emergence of the discipline called visualization. Visualization requires storage of very large image files, e.g., megabytes per image, versus kilobytes per typical alpha numeric file (1000 to 1). The requirements of visualization have not only brought about changes in the design of processors from character to image processing, but also a recognition of the need for larger and faster storage systems.

Another major change affecting all of us is the improvement in data communications; not just satellite links, but network buses which handle high data rates with packet switching protocols for peer-to-peer and client/server communication. Using fiber optics, data of various bandwidths can be digitized, buffered, and transmitted in packets over long distances.¹ Network buses will not only facilitate more distributed computing, but also the

implementation of more and larger mass storage systems.

A recent issue of Business Week lists IBM as the number one U.S. Company in terms of market value.² IBM's sales of magnetic storage products (disks and tape drives) are said to exceed the sales of their mainframes. Annual sales of magnetic recording products now exceed \$50 billion and can be expected to grow.³ Computer manufacturers probably sell more disks than tape drives, but the tape market itself is very large. A recent Seagate ad says that "Seagate shipped over seven million disk drives alone last year (1990)."⁴ The reason for including these impressive statistics here is to help dispel any perceived notion that magnetic recording is out of date. Other new technologies such as optical recording have their place, but this paper will demonstrate the advantages of magnetic tape for large mass storage systems.

For those of us in the magnetic tape industry, we have difficulty hearing words like "peripheral" and seeing pyramid shaped icons with mainframe storage at the top, then solid state disk, then magnetic disk, then optical disk and finally magnetic tape (Figure 1). After all, what media stores all of the permanent data for long periods of time? Well, if we were processing time card charges, point of sale transactions, inventory movement, or some other batch process requiring only short term back-up, we would view the whole process the traditional way. But, if we were processing terabytes of sensor data daily and storing it for years, we would view things differently. We would want a storage technology that is very fast (e.g., HIPPI rates), uses inexpensive media, has high data integrity, uses time proven technology, and standard media and storage devices. Perhaps the stack of storage technology would be the same, only in our view the central memory and direct access storage are only little

bumps on the large tape storage archives (see Figure 2).

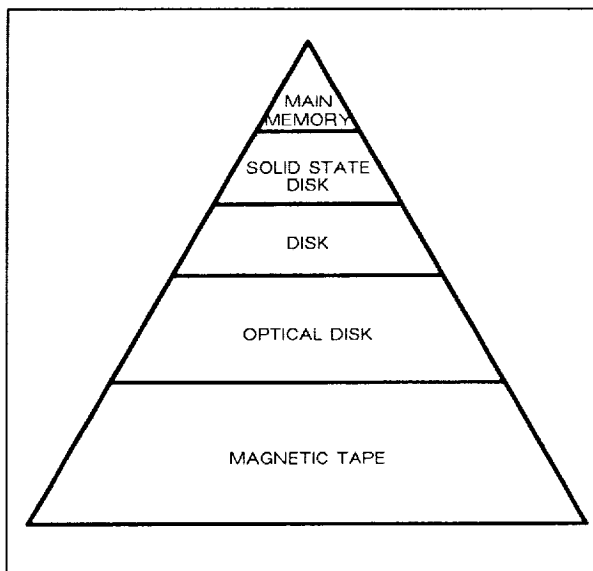


Figure 1. Typical ICON Representation of a Traditional Storage Hierarchy

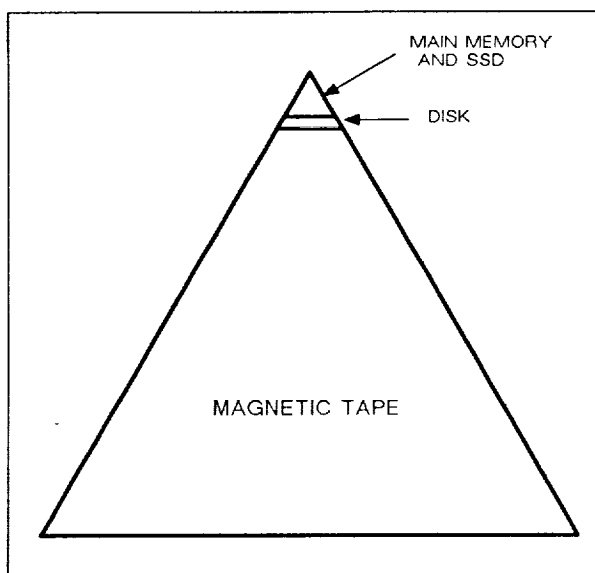


Figure 2. Storage Hierarchy for Large Mass Storage System

Humor aside, a serious attempt will be made to present magnetic tape storage in the context of the conference, i.e., Mass Storage Systems for Space and Earth Science Applications. At first, the writer was tempted to deal with magnetic recording technology from the standpoint of design, but on reflection it was decided to cover only those issues deemed most important to the audience. For those who want to delve more deeply into the design of digital magnetic data recorders, several sources of technical data are recommended. They are 1) NASA Reference Publication 1111, "High Density Digital Recording," prepared by THIC, September 1985, available from the Superintendent of Documents; 2) Magnetic Recording Handbook edited by C. Dennis Mee and Eric D. Daniel and published by McGraw Hill, 1989; 3) Magnetic Recording Handbook edited by Marvin Camras and published by Van Nostrand Reinhold, 1988; 3) The complete handbook of Magnetic Recording by Finn Jorgensen and published by TAB BOOKS INC., 1986

Other valuable sources of data are found in two IEEE publications: (1) November 1986 proceeding with a special section on Magnetic Information Storage Technology edited by Mark Kryder, Professor of Electrical and Computer Engineering and Director of the Magnetics Technology Center of Carnegie Mellon University of Pittsburgh, Pennsylvania. (2) June 1990 special issue on Magnetics incorporating an invited paper by John C. Mallinson, Fellow IEEE and Director of the Center for Magnetics Research, University of California at San Diego. The paper is entitled "Achievements in Rotary Head Magnetic Recording."

TECHNICAL PERSPECTIVE

There is a lack of public understanding, and even a lack of understanding among electrical engineers, concerning the importance of magnetic storage technology. Which came first, computers or magnetic storage? Of course magnetic storage came first, making computers possible.⁵

Magnetic tape recorders record data differently than semiconductor memories. Semiconductor random-access memories and logic devices store 1s and 0s by controlling the state of semiconductor devices. Power for storage and readout of the devices is supplied by the memory power supply or battery back-up.

In digital magnetic recording, bits are stored by creating magnetic domains in moving magnetic media. To store strings of encoded 1s and 0s in magnetic tape, the tape must be moved past a magnetic recording head. Magnetic domains are created in the thin layer of ferromagnetic particles in the surface of the tape by the field of the drive current. Similarly, encoded 1s and 0s are read from the tape by sensing the rate of change of the magnetic flux. Domains are created in the tape at discrete intervals whose lengths are inversely proportional to the recording data rate and the speed of the media past the head.

The two most common magnetic recording techniques are illustrated in Figure 3. In longitudinal recording, modern machines typically have 9, 18, 28, 36, 42, or 84 tracks. Contemporary helical-scan recorders achieve higher track densities and higher bit densities, i.e., flux reversals per unit length of track. A rotary scan technology similar to helical scan is transverse scanning wherein the tracks are orthogonal to the direction of tape travel. This method is mentioned briefly in the next section.

Because the magnetic recording process seems complicated when its compared to semiconductor memory, why is it used? The answer is very low cost and very high storage density at high data rates. As an example, the ANSI ID-1 Helical Scan Recorder stores data on large tape cassettes at a media cost of less than two tenths of a cent per megabyte and at a

density of 625 megabytes per cubic inch with data rates up to 50 MBytes per second.

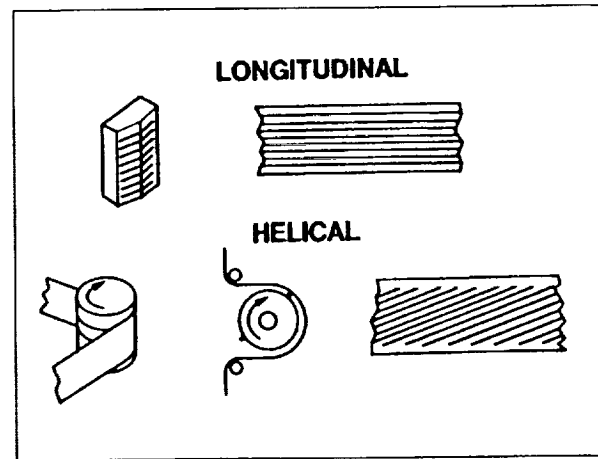


Figure 3. Magnetic Tape Recording Techniques

The energy identifying every bit of storage is stored as a domain of spinning electrons in the particulate media. These domains remain indefinitely unless demagnetized by an external field greater than the coercivity of the tape.

HISTORICAL PERSPECTIVE

Widespread use of magnetic tape recording began in the late 1940's. Since then many recorder configurations have evolved by many manufacturers and a great many still exist with refinements. In the beginning, they were all analog machines recording complex wave forms for instrumentation, audio, and video applications. In 1956, Ampex invented transverse scan rotary recording and launched a product successfully for broadcast video recording.⁶ RCA followed with its transverse scan products in 1958. These machines were called "quadrature" machines because the heads scan across the tape at a quadrature angle with respect to the direction of tape (90°).

The most significant aspect of this early rotary recording technology is that it permitted very high head to tape speeds (over 3000 inches per second) and high track density, hence high video bandwidths and large storage capacity.

Magnetic tape was selected in the early 1950's to transfer data at high speeds to and from computers.⁷ These were 7-track and later 9-track longitudinal recording machines, some of which are still in service. In the 1970's and

1980's, 1/2" tape cartridges and 1/4" tape cassettes were introduced for longitudinal data recording. DATATAPE introduced its System 600, 84 channel longitudinal digital data recorder, in 1980. The System 600 machines are capable of selectable record and reproduce rates from 6.25 to 56 megabytes per second.

The earliest reference found in the literature regarding computer tape drives using rotary heads to achieve high information densities on tape (Damron, et al, 1968) is on p. 6 of chapter 4, referenced above.⁷ Starting in the early 1980's, DATATAPE began producing a number of high data rate, high storage capacity helical scan recorders for DOD. Three of these machines are pictured in Figure 4. All three are still being manufactured. Please note that the RDCR-331 is a mass storage system with automated cassette handling.

Two other versions of the RDCR have been manufactured in quantity. The first uses an NEC jukebox with 3M tape cassettes in the shape of D-1 tape cassettes, but with 1" wide tape instead of 19mm (3/4") tape. The second version produced cooperatively for Masstor Corporation is being sold in commercial mass storage systems.

CONTEMPORARY DIGITAL DATA RECORDERS

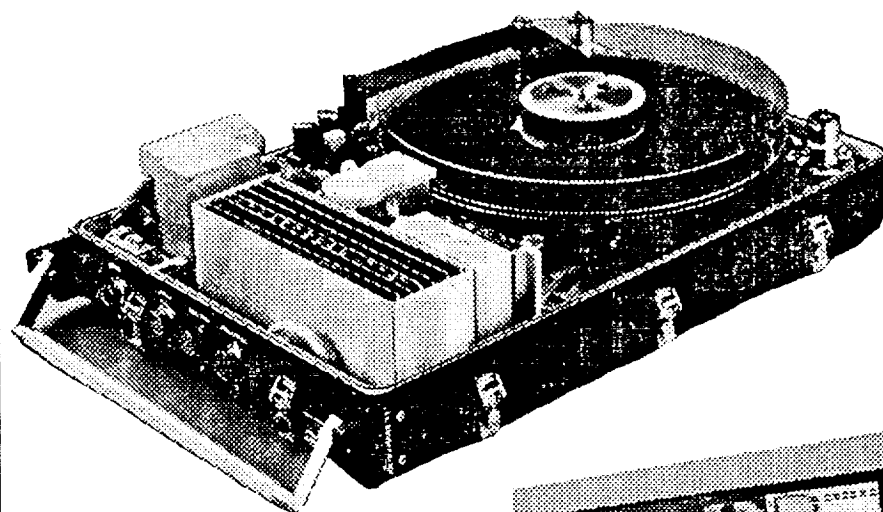
By all odds, the high volume digital data recorders of recent note have been the 3480/3490, 1/2" machines manufactured by IBM, Storage Technology and others. These longitudinal machines operate with a Federal Information Processing Standard -60 (FIPS-60) Interface. Their data rate, 3-4 megabytes per second, and their modest cartridge storage capacity (200 plus megabytes) have been adequate for contemporary mainframe storage. These recorders, manufactured in the thousands, if not hundreds of thousands, sell for reasonable prices, in the range of \$10,000 to \$20,000 each. They are also available in large automated storage silos, e.g., Storage Technology Corporation's ACL-4000 which stores over 1 terabyte.

Also of particular note has been the success of 8mm helical scan tape recorders manufactured

by Exabyte Corporation. Sold by many system integrators and re-sellers for small computer disk back-up applications, these machines are based on the use of a Sony 8mm video tape drive with Exabyte developed digital data electronics. Sold with Small Computer System Interface (SCSI), these machines operate at a burst data rate of 1.5 megabytes per second and store 2.3 gigabytes. A higher density version of this machine is available with 1638 tracks per inch (tpi) versus 819 tracks per inch and stores 5 gigabytes. Both Exabyte machines use metal particle tape. Other value-added suppliers of Exabyte Recorders include Megatape and Summus Corporation. Summus also sells the Exabyte machines with automated carousels. Both of these suppliers offer standard peripheral interfaces.

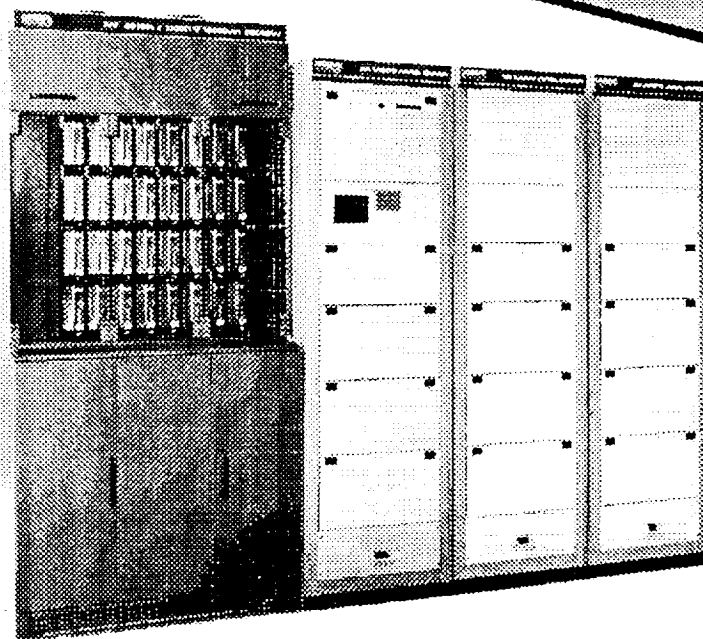
Some additional contemporary machines deserve mention. The first is the Metrum VLDS 1/2" Helical Scan tape cartridge machine based on a VHS product. In a dual channel version it is capable of a 3 megabytes per second data rate and in a single channel version 1.5 megabytes per second. In a read-after-write version, the dual channel machine is reported to be capable of achieving a corrected bit error rate (CBER) of 1 error in 10^{12} bits with a transfer rate of 1 megabytes per second. This machine is being produced in volume. It is available with an automatic cartridge changer which holds up to 600 cartridges. The machine is available with TTL, SCSI-1, VME, and VAX DRB-32 interfaces. The machines use cobalt doped gamma ferric oxide tape on standard 5.2 gigabyte cartridges, with 10 gigabyte cartridges available. The recording format on tape is proprietary to Metrum.

A higher performance transverse scan tape cartridge machine is the AMPEX DCRSi which operates at 13.4 megabytes per second and stores 47.5 gigabytes per cartridge on cobalt doped gamma ferric oxide 1" wide tape. The machine has a 96 megabyte buffer which allows read or write at any rate up to 13.4 megabytes per second. The machine is available with either a serial or 8 bit parallel TTL interface and an RS-232 or RS-422 control and status interface. Several hundred machines are in service. The tape format is Ampex proprietary.



DDR-100
(2 Terabits per reel)
250 GBytes/Reel
44 KBIPI
Track Pitch 1.6 mils

MDR
(5 Terabits per reel)
625 GBytes/Reel
44 KBIPI
Track Pitch 1.2 mils



RDCR-331
(7.6 Terabits per JB)
950 GBytes/Jukebox
Track Pitch 1.2 mils

Figure 4. Recent DATATAPE Helical Scan Digital Data Recorders

Not included in this paper are detailed statistics on a series of 4mm helical scan tape cartridge machines which are based on DAT (digital audio tape) consumer products. Generally speaking, they have transfer rates of 0.19 megabytes per second, read-after-write, storage capacity of 1.25 gigabytes per cartridge, CBER < 1 error in 10^{15} bits, and SCSI-1 and PERTEC interfaces. The machines use 1400 Oersted metal particle tape available from a number of manufacturers. Systems are said to be available from Archive, Caliper, Gigatape, Gigatrend, Hewlett Packard, Hitachi, Sony, Wangtek, Wang Dat and others. There are two proposed standards. The most interesting aspect of these machines is their high data density: 61000 bpi with 535 micro-inch track spacing (1869 tpi).

Another helical scan storage system of note is the Masstor M-960 Cart Machine which operates with the Masstor M-1000 storage Library. These machines are 2 terabyte libraries with FIPS-60 Interfaces and are operated on-line for large enterprises.

HIGH PERFORMANCE 19MM HELICAL SCAN DIGITAL DATA RECORDERS

Several years ago, an industry and government working group was formed to generate a standard for a new 19mm digital data recorder (ID-1). The ID-1 standard would permit D-1 Digital Broadcast Tape Drives to be used with changes to the tape footprint and with new digital data electronics. The data electronics would incorporate Reed Solomon RS 4/5 Error Correction Circuitry and other desirable attributes, e.g., Scan Block ID and longitudinal search and annotation tracks. The machines would be available from several manufacturers and tape cassettes recorded on one manufacturer's machine could be reproduced on another's. This would avoid or at least mitigate the problem anticipated by some "that eternal copying from one format to the next will be necessary".⁸ (If, in fact, some images may be required to be stored for very long periods, photographic film may be the media of choice, e.g., witness civil war photographs.)

The ID-1 Working Group included representatives from the following organizations (see Table 1):

Ampex
Bow Industries
Datatape Incorporated
Fairchild Weston (now Loral)
Honeywell (now Metrum)
Odetics
Memorex
RCA (now GE)
Sandia
Schlumberger
Sony
U.S. Government (NASA and DOD)

Table 1. ID-1 Standard Working Group

The dedicated effort of many individuals over an extended period of time resulted in the approval of the Standard by the American National Standards Institute, Inc., on December 7, 1989. It is identified as ANSI X3.175-1990, American National Standard for Information Systems - 19mm type ID-1 Recorded Instrumentation - Digital Cassette Tape Format; Secretariat, Computer and Business Equipment Manufacturers Association.

In the last two years, new ID-1 products have been announced by most of the companies represented on the working group. They include GE, Metrum, Loral, Sony, and Datatape. The cost to bring such new products to market is estimated to be in the millions of dollars, if not tens of millions of dollars.

At first the writer had planned to tabulate the performance values for the products offered, but on reflection, it was recognized that a mistake might be harmful. At any rate, the data recorded by one manufacturer can be reproduced by any other, right? Yes, we expect it can, but the data might not be useable, if the data has been compressed or encrypted, and the reproducer does not have the equivalent decompression or decryption equipment. The same applies to the use of external buffers and/or enhanced interleaving/error correction techniques. Because of this, an ANSI Working Group has been formed to draft a standard for the use of ID-1 machines modified to meet computer peripheral recorder requirements and meetings are in process.

It should be noted that the ANSI ID-1 standard does not prescribe data rate for recording or reproduction. Because of speed scaling, the foot print on tape is the same regardless of data rate, within limits. Accordingly, some manufacturers are offering machines with lower maximum data rates.

In the meantime, Datatape has announced the availability of a Variable Rate Buffer which can be used with the Datatape DCTR-LP series machines to provide read-after-write and rewrite capability with rates up to 50 megabytes per second using standard computer control and status interfaces with ECL data channels. The DCTR-LP machines can also be used with a CHI Systems interface to operate over a HIPPI channel with IPI-3 protocol. CHI Systems has also announced the availability of its IPI-3 and VME bus interfaces with other manufacturer's ID-1 products. The block diagram of CHI Systems HIPPI-DCTR-LP interface is shown in Figure 5.

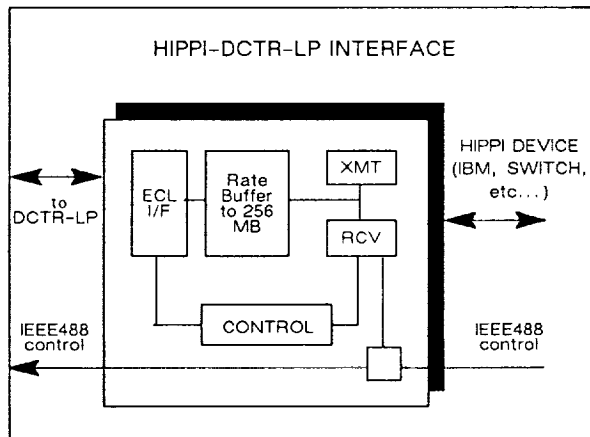


Figure 5. Chi Systems DCTR-LP ID-1 Interface

The Datatape Variable Rate Buffer has up to 448 megabytes of dynamic ram in order to buffer recorder data at rates up to 50 megabytes per second and to provide read-after-write and rewrite.

Some of the features of the Datatape ID-1 machine are illustrated in the following paragraphs as generally representative of all ANSI standard ID-1 machines. It will be noted that some of the illustrations include characteristics representative of MIL-STD-2179,

e.g., zero azimuth tracks which are available from Datatape and others, but, generally speaking, this presentation is intended to highlight ID-1 azimuth recording tracks in which alternate scans are recorded with the azimuth shifted fifteen degrees; first in one direction, then the other. While azimuth recording may cause a slight reduction in SNR when exactly on track, the advantage achieved in raising tracking tolerances is considered advantageous for all concerned and is in keeping with the ANSI X3.175 standard.

A comparison of ferro fluid pictures of 0 degrees azimuth and ± 15 degrees azimuth recording is shown in Figure 6.

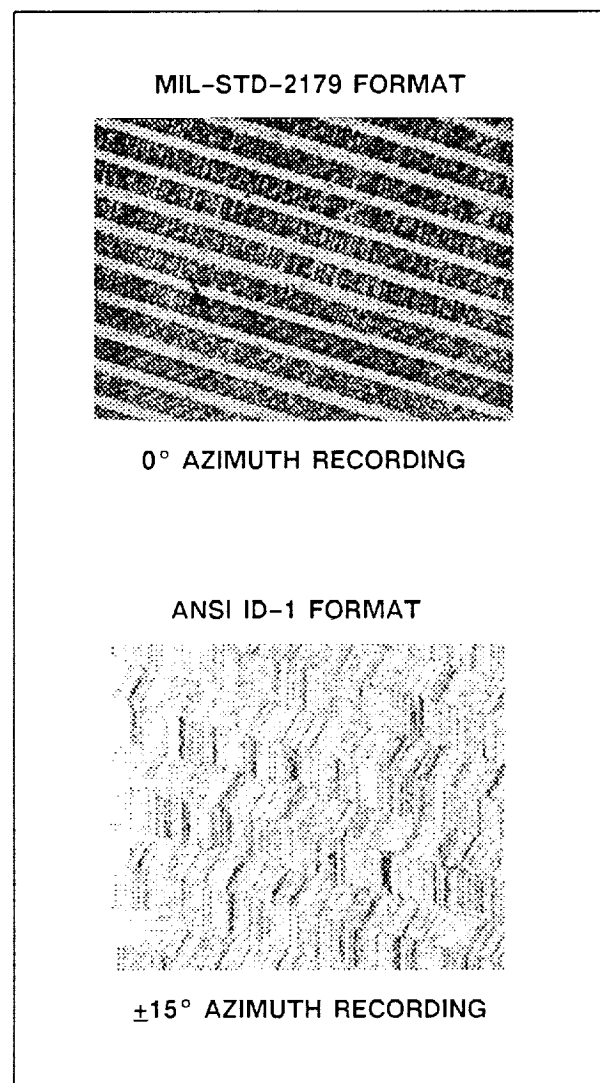


Figure 6. Azimuth Recording

Performance parameters of MIL-STD-2179/ANSI ID-1 Data Recorders are shown in Figure 7.

To provide the reader with an appreciation for the head-to-tape interface for an ANSI ID-1

recorder, the tape format and the Datatape DCTR-LP series scanner is shown in Figure 8.

The thing to note is that at 10,380 RPM the head is literally moving at over 90 miles per hour along a track which is narrower than the diameter of an average human hair.

	SMPTE D-1 DTTR	400 Mbit/s RECORDER	300 Mbit/s RECORDER	200 Mbit/s RECORDER
DRUM DIAMETER	2.95 IN	2.95 IN	2.95 IN	2.95 IN
ACTIVE WRAP ANGLE	257°	256°	257°	258°
AZIMUTH	0°	+15° OR 0°	+15° OR 0°	+15° OR 0°
TRACK PITCH	1.8 MIL (45 µm)	1.8 MIL (45 µm)	1.8 MIL (45 µm)	1.8 MIL (45 µm)
SCANNER RATE	149.9 R/S	173.1 R/S	173.1 R/S	173.1 R/S
NUMBER OF HEADS	4	8	6	4
TAPE VELOCITY	11.3 IN/S	26.1 IN/S	19.5 IN/S	13 IN/S
HEAD-TO-TAPE SPEED	1401 IN/S	1632 IN/S	1625 IN/S	1619 IN/S
USER DATA RATE	179 Mbit/s	400 Mbit/s	300 Mbit/s	200 Mbit/s
TAPE DATA RATE	225.1 Mbit/s	533 Mbit/s	400 Mbit/s	265 Mbit/s
INSTANTANEOUS HEAD DATA RATE	78.7 Mbit/s	93.9 Mbit/s	93.5 Mbit/s	92.6 Mbit/s
LINEAR DENSITY	56 Kbit/IN	57 Kbit/IN	57 Kbit/IN	57 Kbit/IN
SPEED/RATE RANGE	SINGLE SPEED	8:1	8:1	8:1
CHANNEL CODE	RANDOMIZED	8/9 DC FREE CODE FOR VARIABLE SPEED. ANSI- ID-1 ALSO INCLUDES RANDOMIZING		
BER WITH ERROR CORRECTION	1X10 ⁻⁶ W/ECC	1X10 ⁻¹⁰ W/ECC (R/S)	1X10 ⁻¹⁰ W/ECC (R/S)	1X10 ⁻¹⁰ W/ECC (R/S)

Figure 7. MIL-STD-2179/ANSI ID-1 Data Recorders

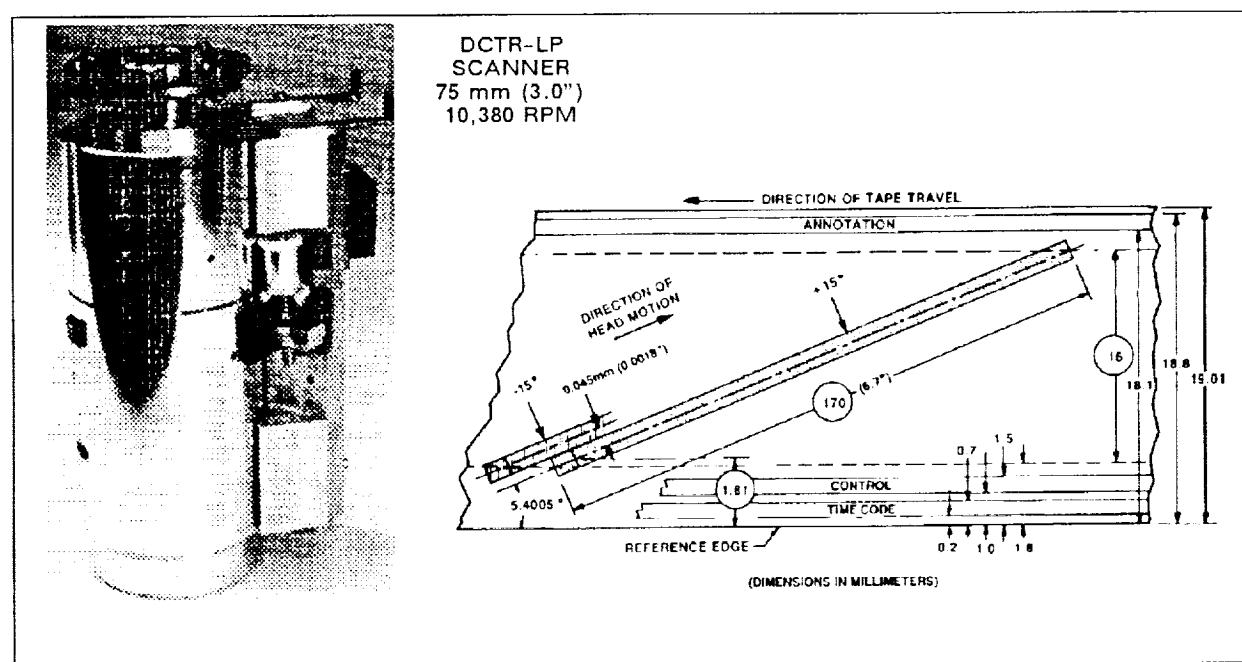


Figure 8. Tape Format and DCTR-LP Scanner

Perhaps more remarkable is the nature of the head itself shown in Figure 9. The head gap length is only 0.3 micro-meters. This is one half the wavelength of visible light (average 0.6 μm). Therefore, since the gap length is shorter than the wavelength of light, the headgap cannot be viewed directly with an optical microscope. It is viewed only by a scanning electron microscope (SEM). Is it any wonder then that optical 3.5" and 5.25" disks have linear recording densities of only 31.8 kilobits per inch (kbpi) compared to 57 kbpi for the ID-1 machine, 75 kbpi for the 8mm Helical Scan Machine, and 38 kbpi for the 3480.⁹ The point to be made in comparing the linear density of the magnetic recorders to the optical disks is not to say that the magnetic recorder is superior, per se. They are in the same general ball park with regard to linear track density, but tape has other obvious advantages like data rate, capacity, and cost per megabyte.

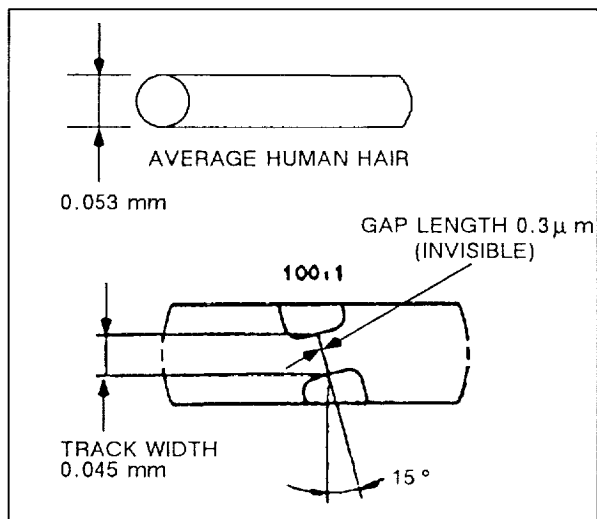


Figure 9. ANSI ID-1 Record Head used on DCTR-LP Series Recorders

The rotary headwheel for the DCTR-LP series ID-1 machines is shown in Figure 10. There are 16 sections on the rotary transformer making it possible to use up to 8 record heads and 8 reproduce heads to achieve a 50 megabytes per second data rate. It should be noted that to go from 25 megabytes per second to 50 megabytes per second it is not necessary to push the record and reproduce circuitry for

higher frequency performance; it is merely necessary to use more circuits, i.e., head drivers, pre-amps, equalizers, etc.

The tape cassettes used in the ANSI ID-1 machines are shown in Figure 11. Please note that 13 μm tape is listed as well as 16 μm tape even though, at this time, only 16 μm thick tape (0.4") tape is being used in ID-1 machines.

The Datatape ID-1 Laboratory Machine, the DCTR-LP series, is illustrated in Figure 12. These machines are designed for 19-inch rack mounting either as one unit 30 inches deep or as two units side by side (see Figure 12). This laboratory version of the Datatape DCTR uses a tape drive developed by BTS in Germany together with the data electronics unit developed at Datatape in Pasadena.

An airborne version of the ID-1 machine is shown in Figure 13. It has been designed to meet MIL-E-5400 environments. This machine is manufactured in Pasadena, except for the scanner which is made in the Datatape, Santa Clara, California Facility.

The ID-1 machines designed for digital data collection and for mass storage with data buffering, represent a broad industry push to satisfy a generation of digital data storage requirements. A conservative approach has been taken in advancing the ID-1 designs as noted below:

1. Track Spacing: 0.045mm (0.0018")
The tracks are contiguous on tape (no lost space) and the tracks are generous, both in width and bit density. This is important when using the machine for recording or reproducing short bursts where longer time intervals are not available, e.g., to lock-up head tracking servos.
2. The 850 Oersted tape also represents a conservative design approach. It uses cobalt treated gamma ferric oxide tape which has been proved in over 20 years of service.

3. High data rates and storage capacities are achievable with current designs, i.e., 50 megabytes per second vs 15 megabytes per second for the D-2 based digital data recorder discussed briefly below.
4. With multiple suppliers of equipment designed to standards, it is highly unlikely that a user will get caught with an archive and no way to reproduce the data.

Ampex Corporation has announced a new R-90 product based on using their VPR-300 D-2 Rotary Digital Video Broadcast Machine for digital data storage applications. Some of its published features are listed in Table 2 below.

- Helical scan recorder
- 15 Megabyte per second transfer rate
- IPI-3 interface
- 300 inch per second shuttle speed
- 150 inch per second search speed
- Accepts three sizes of cassettes
 - Small cassette holds 25 gigabytes
 - Medium cassette holds 75 gigabytes
 - Large cassette holds 165 gigabytes
- Less than one permanent read error per 10^{12} bytes read
 - Interleaved Reed-Solomon ECC
 - Read-after-write with selectable rewrite
 - Automatic reread for soft errors

Table 2. Ampex R-90 Recorder Features

Ampex has initiated an effort to establish an ANSI standard for the R-90 machine. So far, to the writers knowledge, no other manufacturer has announced plans to manufacture a compatible machine, but all of the ID-1 manufacturers will be represented on the working group.

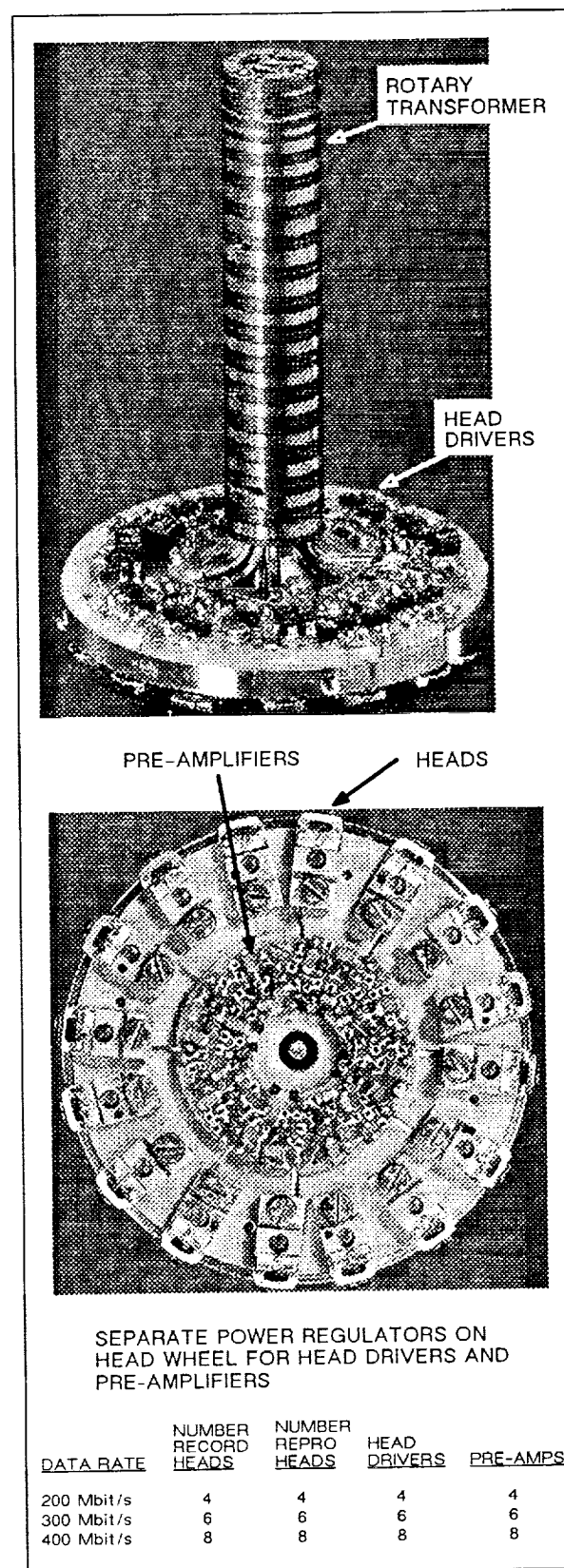


Figure 10. Rotary Headwheel for the DCTR-LP Series ID-1 Machines

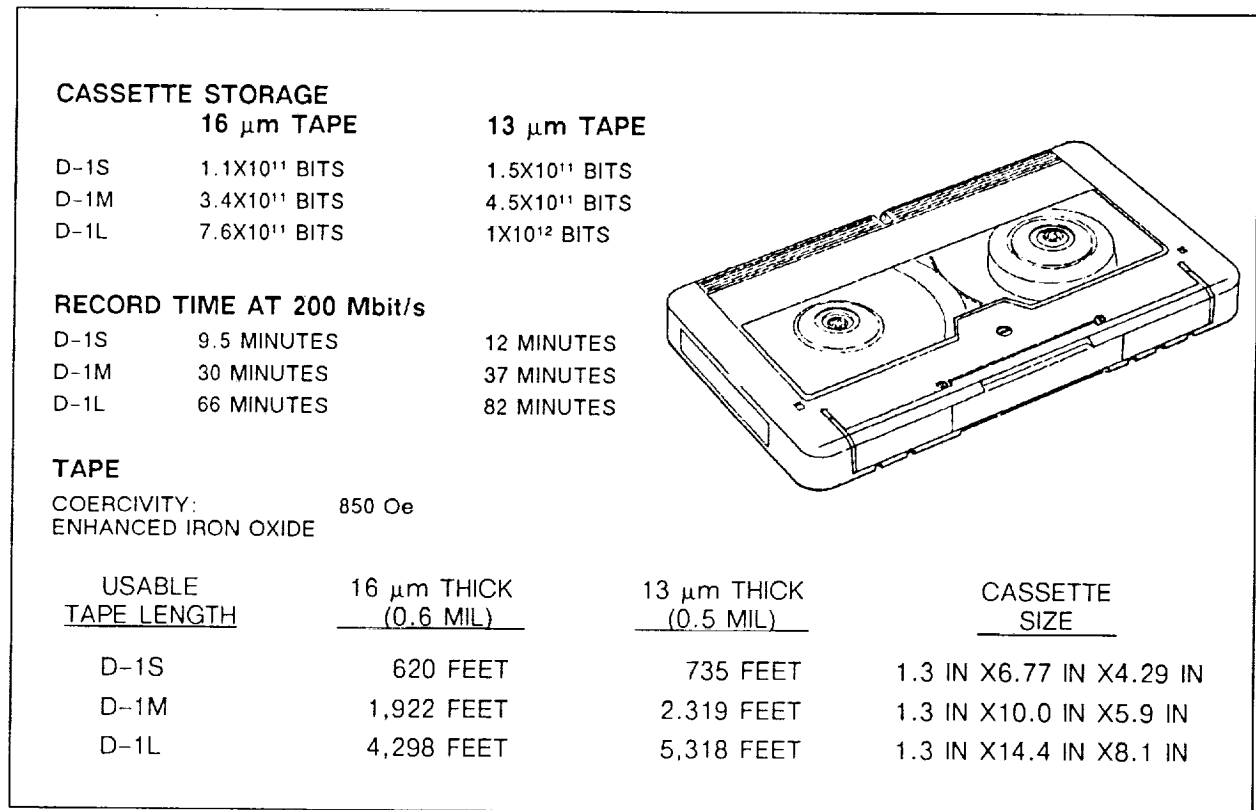


Figure 11. ANSI ID-1 Tape Cassettes

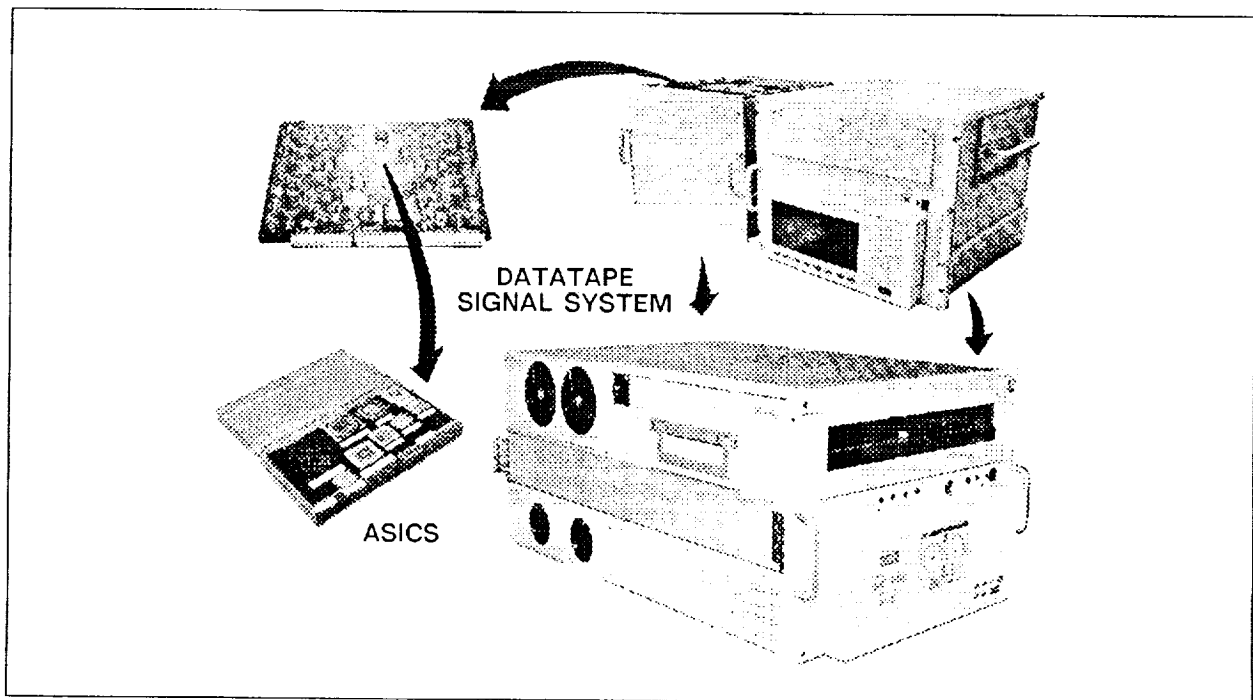


Figure 12. Datatape ID-1 Laboratory Machine, the DCTR-LP Series

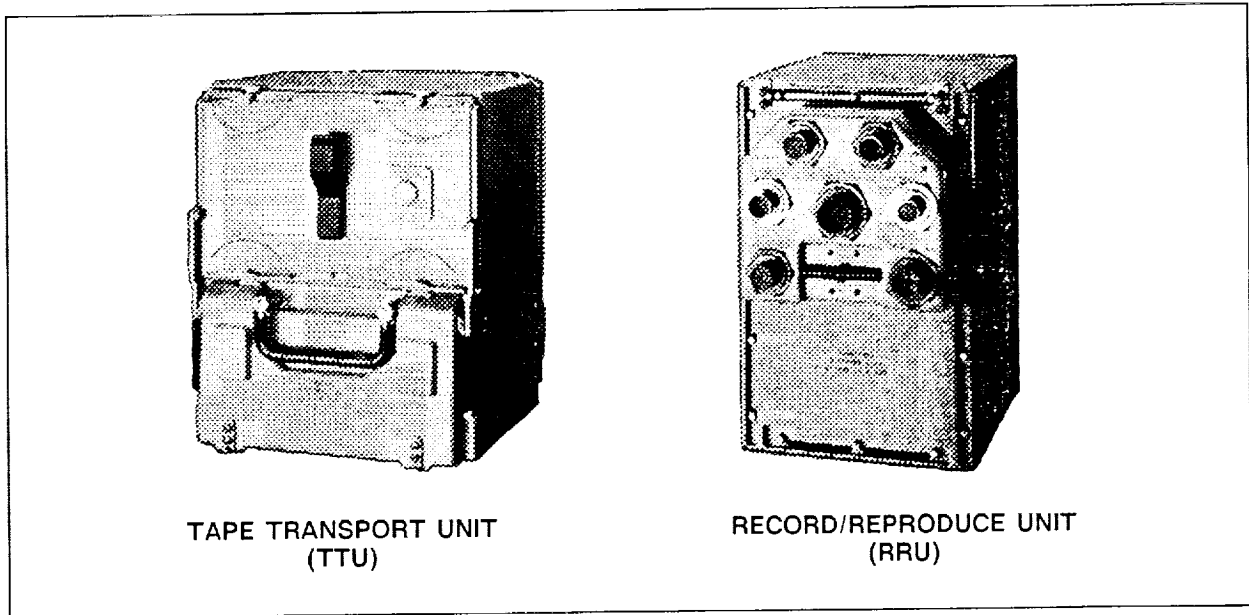


Figure 13. Airborne ID-1 Machine, the DCTR-A120

FUTURE DEVELOPMENT

New and exciting developments are evolving from research in magnetic tape recording, particularly with respect to tape and magnetic heads.

It seems that everywhere you turn there is a new recorder product using metal particle 1400 or 1500 Oersted tape, e.g., 4mm DAT, 8mm, 1/2" M-2/D-3, and 19mm D-2. Since the magnetic field required to magnetically saturate the metal particle tape is higher than for conventional tapes, new heads are required. Ferrite heads saturate at lower field strengths. Accordingly, new heads such as metal-in-gap heads have been widely used.

DATATAPE and Kodak Research Laboratories have developed new metal laminated heads which work very well with higher coercivity metal particle tape. These heads have permitted an increase of per channel data rates from 50 to 200 megabits per channel. This means that a machine with 8-record and 8-reproduce channels will be able to operate at 1,600 megabits per second (200 Megabytes per second), a rate substantially higher than that required for HDTV.

We are following closely, on-going development of barium-ferrite as well as metal particle tape,

and we look forward to the panel discussion at this conference.

SUMMARY

The machines required for storage and retrieval of bit files of any length are more demanding than those designed for streaming disk back-up or continuous video recording. Data integrity and reliability are absolute requirements. In addition, data recorded on a cassette in one machine must be reproducible in another machine.

To achieve these goals, great care must be exercised in the design of tape guidance and servo coupling between the scanner control track, the tape drive capstan and reel motors; but most importantly, designs must be generous where it also counts, i.e., track width and bit density.

All tape recorders experience errors when writing and reading data to and from magnetic tape. The approach must be to minimize the factors which contribute to errors (tape quality and tracking) and to incorporate robust error correction means to improve performance (interleaving, coding, and write-read and rewrite).

The ANSI ID-1/DD-1 machines incorporate the desirable features for current generation high speed, high capacity digital data recorders. The next year should prove to be very exciting as the new generation of 19-mm machines enter service in data storage applications.

ACKNOWLEDGMENTS

The writer is indebted to the following individuals who contributed to the editorial review of this paper: Hank Tobin, Norris Huse, Ray Aires, Bill Bullers, Leo Van Lahr, Mike Trcka and Joe Trost. Credit is also due to Joyce Smulo and Eve FitzGerald for generating the reproducible copy on a very tight schedule.

REFERENCES

- 1 Mukherjee, Biswanath, "Integrated Voice-Data Communications over High-Speed Fiber Optic Networks," Computer, IEEE Computer Society, February 1991, pp. 49-57.
- 2 The Top 1000 U.S. Companies, Business Week, 1991 Special Issue, 1991, pp. 100-208
- 3 Wood, Roger, "Denser Magnetic Memory," IEEE Spectrum, May 1990, pp. 32-38
- 4 Advertisement by Seagate, Computer Technology Review, May 1991, pp. 24-25.
- 5 Kryder, Mark R. and Seger, Paul J., "The special section on Magnetic Information Storage Technology," Guest Editor, Proceedings of IEEE, Vol 74, No. 11, Nov. 1986, pp. 1475-1476.
- 6 Mallinson, John C., "Achievements in Rotary Head Magnetic Recording," Proceedings of the IEEE, Vol. 78, No. 6, 1990, pp. 1004-1006.
- 7 Cannon, Max R. and Seger, Paul J., "Data Storage on Tape." Magnetic Recording Handbook, C. Dennis Mee and Eric D. Daniel, McGraw-Hill, 1990, pp. 815-882.
- 8 Mallinson, John G., "Magnetic Tape Recording: Archival Considerations," Tenth IEEE Symposium on Mass Storage Systems, IEEE Catalog No. 90CH2844-9, May 1990, pp. 58-60.
- 9 Katzive, Bob, "Product Projections," IIST Workshop VI, Lake Arrowhead, October, 1990, Session VII, p. 2.

MAGNETIC TAPE

HARRISS ROBINSON

DATATAPE INCORPORATED

23 JULY 1991

OVERVIEW

- DATA STORAGE AND PROCESSING ENVIRONMENT
- MAGNETIC TAPE TECHNICAL PERSPECTIVE
- CONTEMPORARY DIGITAL DATA RECORDERS
- HIGH PERFORMANCE 19mm HELICAL SCAN
DIGITAL DATA RECORDERS
- FUTURE DEVELOPMENT
- SUMMARY

HARRISS ROBINSON DATATAPE INCORPORATED

PRECEDING PAGE BLANK NOT FILMED
1-93

DATA STORAGE AND PROCESSING ENVIRONMENT

- **MOVE TO VISUALIZATION REQUIRES**
 - **VERY LARGE FILES FOR IMAGES VS ALPHANUMERIC FILES (1000 TO 1)**
 - **CHANGES IN DESIGN OF PROCESSORS**
- **IMPROVED DATA COMMUNICATIONS THROUGH NETWORK BUSES**
 - HIGH DATA RATES**
 - PACKET SWITCHING PROTOCOLS**
 - FIBER CHANNELS**
 - DISTRIBUTED COMPUTING**
 - MORE AND LARGER MASS STORAGE SYSTEMS**

HARRISS ROBINSON DATATAPE INCORPORATED

2

MAGNETIC RECORDING IS A ROBUST TECHNOLOGY

- **ANNUAL SALES OF MAGNETIC RECORDING PRODUCTS EXCEED \$50 BILLION**
- **IBM IS LARGEST US COMPANY IN TERMS OF MARKET VALUE ***
— \$74,996,000,000 (GM W/EDS AND HAC — \$29,838,000,000)
- **SEAGATE SHIPPED SEVEN MILLION DISK DRIVES IN 1990**
- **SALES OF MAGNETIC TAPE PRODUCTS SOARED IN 1990**
(IBM STORAGE TECHNOLOGY, EXABYTE, ET. AL.)

* Business Week, May 1991

HARRISS ROBINSON DATATAPE INCORPORATED

MAGNETIC TAPE TECHNICAL PERSPECTIVE

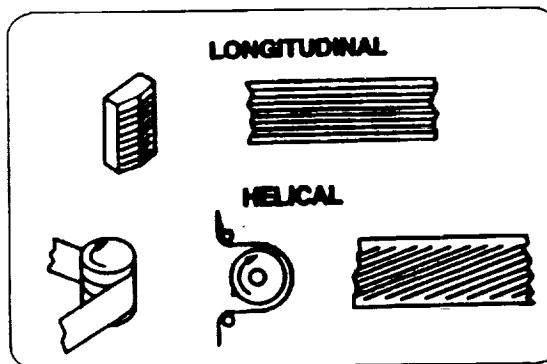
- WHICH CAME FIRST MAGNETIC STORAGE OR COMPUTERS?
- MAGNETIC RECORDERS STORE DATA DIFFERENTLY THAN SEMICONDUCTOR MEMORIES
 - SEMICONDUCTOR MEMORIES STORE 1s AND 0s BY CONTROLLING THE STATE OF SEMICONDUCTOR DEVICES
 - MAGNETIC RECORDERS STORE 1s AND 0s AS DOMAINS OF SPINNING ELECTRONS IN A THIN LAYER OF FERROMAGNETIC PARTICLES

HARRISS ROBINSON DATATAPE INCORPORATED

4

MAGNETIC TAPE TECHNICAL PERSPECTIVE

- COMMON RECORDING TECHNIQUES
 - LONGITUDINAL TRACKS (9, 18, 28, 42, 84 TRACKS)
 - ROTARY SCANNING
 - HELICAL
 - TRANSVERSE



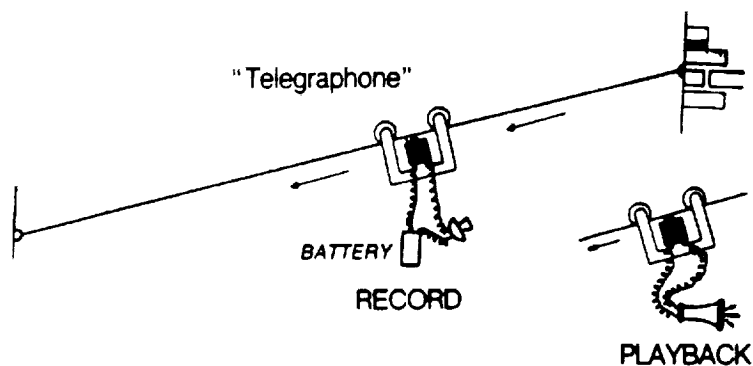
MAGNETIC TAPE RECORDING TECHNIQUES

HARRISS ROBINSON DATATAPE INCORPORATED

5

HISTORICAL PERSPECTIVE

■ FIRST MAGNETIC RECORDING BY PAULSEN 1898



(Magnetic Recording Handbook, by Marvin Camras)

HARRISS ROBINSON DATATAPE INCORPORATED

HISTORICAL PERSPECTIVE (Cont'd)

- LATE 1940s – BEGINNING WIDESPREAD USE OF LONGITUDINAL ANALOG RECORDING OF COMPLEX WAVEFORMS (INSTRUMENTATION, AUDIO, AND VIDEO)
- EARLY 1950s – 7-TRACK AND LATER 9-TRACK LONGITUDINAL RECORDERS SELECTED FOR COMPUTER DIGITAL DATA RECORDING
- 1956 AMPEX INVENTED TRANSVERSE SCAN FOR VIDEO BROADCASTING TV RECORDERS
- 1958 – RCA INTRODUCED TRANSVERSE SCAN VIDEO BROADCAST MACHINES FOR TV BROADCAST
- 1968 – DAMRON, et. al., EARLIEST USE OF ROTARY HEADS TO ACHIEVE HIGH INFORMATION DENSITY FOR COMPUTER DATA (AMPEX)

HARRISS ROBINSON DATATAPE INCORPORATED

HISTORICAL PERSPECTIVE (Cont'd)

- 1970s AND 1980s -
 - 1/4-INCH AND 1/2-INCH TAPE CARTRIDGES WERE INTRODUCED FOR LONGITUDINAL DIGITAL DATA RECORDING
 - DATATAPE INTRODUCED SYSTEM 600, AND 84-TRACK DIGITAL DATA RECORDER WITH SELECTABLE RECORD AND REPRODUCE RATES FROM 6.25 TO 56 MBPS
- 1980s - DATATAPE INTRODUCED SEVERAL HIGH DATA RATE, HIGH STORAGE CAPACITY, HELICAL SCAN DIGITAL RECORDERS FOR DoD. (SEE NEXT VG)
 - THE RDCR IS A MASS STORAGE SYSTEM WITH AUTOMATED CASSETTE HANDLING
 - MODIFIED RDCR TAPE DRIVES ALSO USED WITH NEC JUKEBOX
 - RDCR ALSO USED IN COMMERCIAL MASS STORAGE SYSTEM SOLD BY MASSTOR CORPORATION

HARRISS ROBINSON DATATAPE INCORPORATED

CONTEMPORARY DIGITAL DATA RECORDERS

- 3480/3490 IBM, STORAGE TECHNOLOGY, et. al.
 - LARGEST SALES VOLUME PRODUCT
 - 1/2-INCH LONGITUDINAL TAPE (CARTRIDGE)
 - FIPS-60 INTERFACE
 - DATA RATE: 3-4 MBPS
 - STORAGE CAPACITY: 200 MB
 - SALES PRICE OF TAPE DRIVES: \$10,000 TO \$20,000 EACH; CARTRIDGE APPROXIMATELY \$6 EACH
 - AVAILABLE IN STORAGE TEK™ ACL-4000 AUTOMATED STORAGE SILOS (UP TO 6000 CASSETTES STORING OVER 1 TERABYTE)

HARRISS ROBINSON DATATAPE INCORPORATED

CONTEMPORARY DIGITAL DATA RECORDERS (Cont'd)

■ 8-mm HELICAL SCAN RECORDERS

- MANUFACTURED BY EXABYTE USING SONY 8MM TAPE DRIVE WITH EXABYTE ELECTRONICS
- SOLD THROUGH SYSTEM INTEGRATORS AND RESELLERS FOR SMALL COMPUTER DISK BACK-UP APPLICATIONS
- SCSI INTERFACE
- STANDARD VERSION
 - BURST DATA RATE: 1.5 MBPS
 - STORAGE CAPACITY: 2.3 GB
 - 812 TRACKS PER INCH
- HIGHER DENSITY VERSIONS
 - BURST DATA RATE: 3.0 MBPS
 - STORAGE CAPACITY: 5.0 GB
 - 1638 TRACKS PER INCH
- USES METAL PARTICLE TAPE
- VALUE ADDED SUPPLIERS: MEGATAPE, SUMMUS, et. al.

HARRISS ROBINSON DATATAPE INCORPORATED

10

CONTEMPORARY DIGITAL DATA RECORDERS (Cont'd)

■ 1/2-INCH METRUM VLDS HELICAL SCAN TAPE CARTRIDGE MACHINE

- MANUFACTURED BY METRUM USING A MATSUSHITA VHS TAPE DRIVE
WITH METRUM ELECTRONICS
- SOLD IN VOLUME DIRECTLY BY METRUM
- TTL, SCSI-1, VME, AND VAX DRB INTERFACES
- SINGLE CHANNEL VERSION
 - DATA RATE: 1.5 MBPS
 - CASSETTE STORAGE CAPACITY: 5.2 GB (10 GB AVAILABLE)
- DUAL CHANNEL VERSION
 - DATA RATE: 3 MBPS
 - CASSETTE STORAGE CAPACITY: 5.2 GB (10 GB AVAILABLE)
 - READ-AFTER-WRITE VERSION CBER: 1 ERROR IN 10^{12} BITS
 - COBALT DOPED GAMMA FERRIC OXIDE TAPE
 - AVAILABLE WITH AUTOMATIC CARTRIDGE CHANGER HOLDING UP TO
600 CARTRIDGES

HARRISS ROBINSON DATATAPE INCORPORATED

9

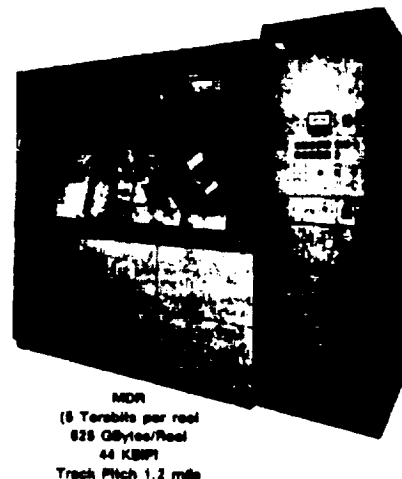
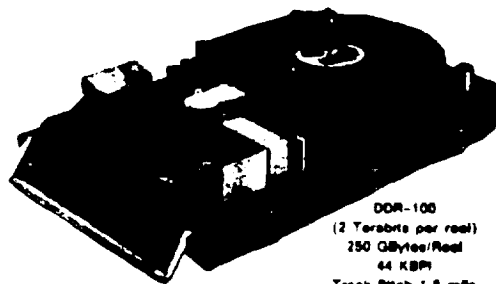
CONTEMPORARY DIGITAL DATA RECORDERS (Cont'd)

- **1-INCH AMPEX DCRSi TRANSVERSE SCAN MACHINE**
 - DATA RATES UP TO 13.4 MBPS USING A 96 MB BUFFER
 - STORES 47.5 GB PER CARTRIDGE
 - AVAILABLE WITH EITHER A SERIAL OR 8-BIT PARALLEL TTL INTERFACE AND RS-232 OR RS-422 CONTROL AND STATUS INTERFACE
 - COBALT DOPED GAMMA FERRIC OXIDE TAPE
 - THE TAPE FORMAT AND TAPE CARTRIDGE ARE AMPEX PROPRIETARY
 - SEVERAL HUNDRED MACHINES ARE IN SERVICE

HARRISS ROBINSON DATATAPE INCORPORATED

12

CONTEMPORARY DIGITAL DATA RECORDERS (Cont'd)



HARRISS ROBINSON DATATAPE INCORPORATED

13

CONTEMPORARY DIGITAL DATA RECORDERS (Cont'd)

■ 4-mm HELICAL SCAN TAPE CARTRIDGE MACHINES

- AVAILABLE FROM: ARCHIVE, CALIPER, GIGATAPE, GIGATREND, HEWLETT PACKARD, HITACHI, SONY, WANGTEK, WANG DAT, et. al.
- TRANSFER RATES OF 0.19 MBPS, READ-AFTER-WRITE CAPABILITY 1.25 GB PER CARTRIDGE, CBER < 1 ERROR IN 10¹⁵ BITS, SCSI-1 AND PERTEC INTERFACES
- TWO PROPOSED STANDARDS
- HIGH DENSITY - 61000 BPI AND 1869 TPI

■ MASSTOR M-960 HELICAL SCAN CART MACHINE

- UP TO 8 M961 HELICAL SCAN STORAGE MODULES @ 110 GB EACH
- M-1000 USER CAPACITY 1,011 GB
- STORES UP TO 6 TERABYTES W/DATA COMPRESSION
- INFORMATION STORAGE AND RETRIEVAL ON LINE, CENTRALIZED, ERASABLE, MAINFRAME ATTACHED, SYSTEM MANAGED STORAGE UNIT WITH APPLICATION SOFTWARE
- SUPPORTS FROM 4 TO 8 IBM HOSTS (4.5 MBPS)

HARRISS ROBINSON DATATAPE INCORPORATED

14

HIGH PERFORMANCE 19-mm HELICAL SCAN DIGITAL TAPE RECORDERS

- APRIL 1986 - INDUSTRY AND GOVERNMENT WORKING GROUPS FORMED TO GENERATE A NEW ANSI STANDARD FOR A 19mm ID-1 INSTRUMENTATION RECORDER
- USE SMPTE STD 19-mm DIGITAL VIDEO TV TAPE DRIVES WITH EACH MFGRS NEWLY DESIGNED DIGITAL DATA ELECTRONICS
 - SCAN TRACKS ON TAPE TO BE CONTINUOUS INSTEAD OF SEGMENTED COMPONENT VIDEO FORMAT
 - USE RS 4/5 ECC, SCAN BLOCK ID, AND LONGITUDINAL SEARCH AND ANNOTATION TRACKS
 - CROSSPLAY OF CASSETTES COULD BE DEMONSTRATED

HARRISS ROBINSON DATATAPE INCORPORATED

15

HIGH PERFORMANCE 19-mm HELICAL SCAN DIGITAL TAPE RECORDERS

■ ID-1 WORKING GROUP INCLUDED REPRESENTATIVES FROM

AMPEX	ODETICS
BOW INDUSTRIES	MEMOREX
DATATAPE INCORPORATED	RCA (NOW GE)
FAIRCHILD WESTON (NOW LORAL)	SANDIA
HONEYWELL (NOW METRUM)	SCHLUMBERGER
	SONY
U.S. GOVERNMENT (NASA AND DoD)	

■ APPROVAL OF NEW STD-7 DECEMBER 1989

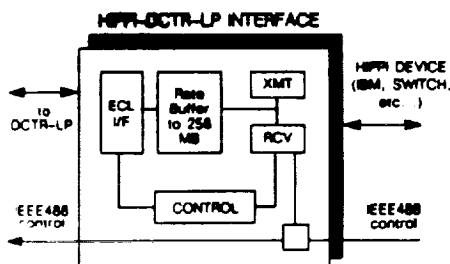
ANSI X 3.175-1990 - 19mm TYPE ID-1 RECORDED INSTRUMENTATION -
DIGITAL CASSETTE FORMAT; SECRETARIAT, COMPUTER AND BUSINESS
EQUIPMENT MANUFACTURERS ASSOCIATION

HARRISS ROBINSON DATATAPE INCORPORATED

16

HIGH PERFORMANCE 19-mm HELICAL SCAN DIGITAL TAPE RECORDERS

- NEW PRODUCTS ANNOUNCED IN LAST TWO YEARS BY
GE, METRUM, LORAL, SONY, AND DATATAPE
- DATATAPE INCORPORATED VARIABLE RATE BUFFER TO OPERATE
UNDER PROTOCOL CONTROL FROM MATCHING ID-1 RECORDER TO
HOST SYSTEM DATA RATES (0 TO 50 MBPS) AND PROVIDE
READ-AFTER-WRITE AND REWRITE
- SYSTEM ALSO PROVIDED BY DATATAPE AND OTHERS WITH
CHI SYSTEMS INTERFACE



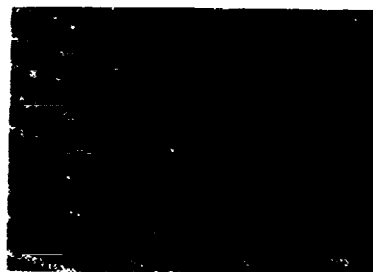
HARRISS ROBINSON DATATAPE INCORPORATED

17

HIGH PERFORMANCE 19-mm HELICAL SCAN DIGITAL TAPE RECORDERS

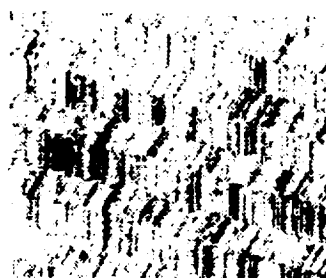
- ANSI X3.175 STANDARD SPECIFIES AZIMUTH RECORDING, BUT
NON-AZIMUTH RECORDERS ARE AVAILABLE FOR MIL-STD-2179

MIL-STD-2179 FORMAT



0° AZIMUTH RECORDING

ANSI ID-1 FORMAT



+15° AZIMUTH RECORDING

HARRISS ROBINSON DATATAPE INCORPORATED

HIGH PERFORMANCE 19-mm HELICAL SCAN DIGITAL TAPE RECORDERS

- PERFORMANCE PARAMETERS OF ANSI ID-1 RECORDERS

	SMPTE D-1 DTTR	400 Mb/s RECORDER	300 Mb/s RECORDER	200 Mb/s RECORDER
DRUM DIAMETER	2.95 IN	2.95 IN	2.95 IN	2.95 IN
ACTIVE WRAP ANGLE	257°	258°	257°	258°
AZIMUTH	0°	+15° OR 0°	+15° OR 0°	+15° OR 0°
TRACK PITCH	1.8 MIL (45 μm)	1.8 MIL (45 μm)	1.8 MIL (45 μm)	1.8 MIL (45 μm)
SCANNER RATE	149.9 R/S	173.1 R/S	173.1 R/S	173.1 R/S
NUMBER OF HEADS	4	8	8	4
TAPE VELOCITY	11.3 IN/S	28.1 IN/S	19.5 IN/S	13 IN/S
HEAD-TO-TAPE SPEED	1401 IN/S	1832 IN/S	1825 IN/S	1619 IN/S
USER DATA RATE	179 MB/s	400 MB/s	300 MB/s	200 MB/s
TAPE DATA RATE	226.1 MB/s	533 MB/s	400 MB/s	285 MB/s
INSTANTANEOUS HEAD DATA RATE	78.7 MB/s	93.9 MB/s	93.6 MB/s	92.6 MB/s
LINEAR DENSITY	56 KB/IN	57 KB/IN	57 KB/IN	57 KB/IN
SPEED/RATE RANGE	SINGLE SPEED	8:1	8:1	8:1
CHANNEL CODE	RANDOMIZED	8/1 DC FREE CODE FOR VARIABLE SPEED. ANSI ID-1 ALSO INCLUDES RANDOMIZING		
BER WITH ERROR CORRECTION	1X10 ⁻⁶ W/ECC	1X10 ⁻¹⁰ W/ECC (R/S)	1X10 ⁻¹⁰ W/ECC (R/S)	1X10 ⁻¹⁰ W/ECC (R/S)

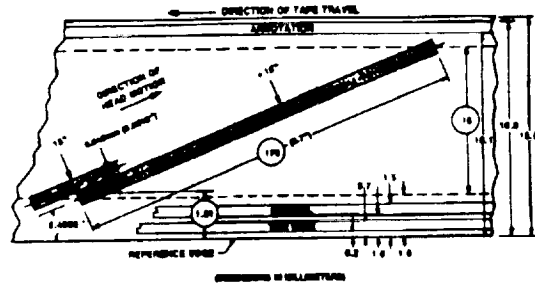
HARRISS ROBINSON DATATAPE INCORPORATED

HIGH PERFORMANCE 19-mm HELICAL SCAN DIGITAL TAPE RECORDERS

ANSI ID-1 TAPE FORMAT AND DATATAPE DCTR-LP SCANNER



DCTR-LP
SCANNER
75 mm
(3.0")
10,380 RPM



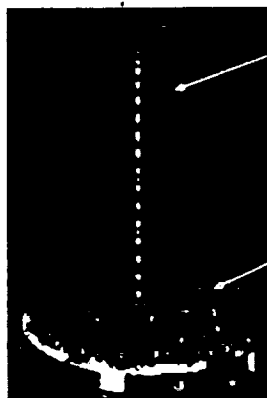
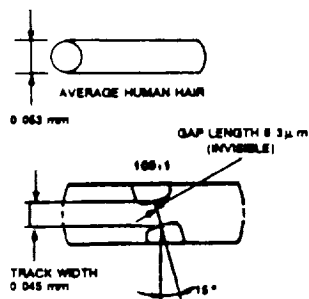
NOTE TOP SPEED OF SCANNER AT 10,380 RPM THE HEADS ARE
LITERALLY MOVING AT OVER 90 MPH ALONG A TRACK THAT IS
NARROWER THAN A HUMAN HAIR

HARRISS ROBINSON DATATAPE INCORPORATED

20

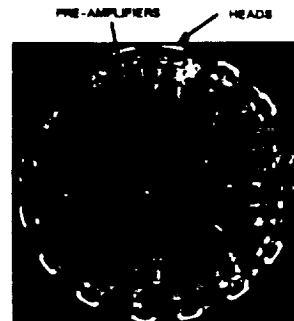
HIGH PERFORMANCE 19-mm HELICAL SCAN DIGITAL TAPE RECORDERS

ANSI ID-1 RECORD HEAD USED ON DCTR-LP SERIES RECORDERS



ROTARY
TRANSFORMER

HEAD
DRIVERS



PRE-AMPLIFIERS HEADS

SEPARATE POWER REGULATORS ON
HEAD WHEEL FOR HEAD DRIVERS AND
PRE-AMPLIFIERS

DATATAPE	HARRISS RECORD HEADS	HARRISS REPRO HEADS	HEAD DRIVERS	PRE-AMPS
200 MBH/70	1	1	1	1
300 MBH/70	1	1	1	1
400 MBH/70	1	1	1	1

HARRISS ROBINSON DATATAPE INCORPORATED

21

AVERAGE DIAMETER OF A HUMAN HAIR* (MILS)

DONOR	M/F (SEX)	MEASURED BY			MEAN
		J. UDELL	L. VAIL	D. HOSKINS	
1. J. UDELL	(M)	2.0	2.0	2.0	2.0
2. H. ROBINSON	(M)	1.4	1.3	1.3	1.4
3. L. VAIL	(M)	2.3	2.3	2.3	2.3
4. M. VIERIA	(M)	2.7	3.0	2.7	2.8
5. K. STOFFERS	(F)	2.3	2.2	2.1	2.2

AVERAGE THICKNESS 2.1
(0.053 mm)

*DATA TAKEN AT DATATAPE 4/25/90

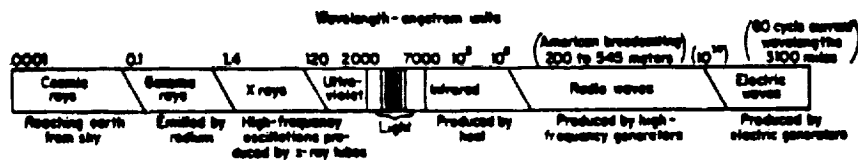
HARRISS ROBINSON DATATAPE INCORPORATED

22

TERMINOLOGY / CONVERSION FACTORS

WORD	ABREV.	BYTES	BITS	NO. BYTES
BYTE	B	1	8	1^0
KILOBYTE	KB	1,000	8,000	10^3
MEGABYTE	MB	1,000,000	8,000,000	10^6
GIGABYTE	GB	etc	etc	10^9
TERABYTE	TB	↓	↓	10^{12}
PETABYTE	PB			10^{15}
EXABYTE	EB			10^{18}
HARIBYTE	HB			10^{21}

(PROPOSED HERE TODAY)



HARRISS ROBINSON DATATAPE INCORPORATED

23

HIGH PERFORMANCE 19-mm HELICAL SCAN DIGITAL TAPE RECORDERS

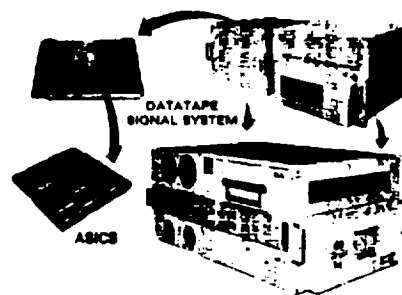
■ ANSI ID-1 TAPE CASSETTE



USABLE TAPE LENGTH	18 μ m THICK (0.6 MIL)	13 μ m THICK (0.5 MIL)	CASSETTE SIZE
D-1S	620 FEET	735 FEET	1.3 IN X 6.77 IN X 4.29 IN
D-1M	1,922 FEET	2,319 FEET	1.3 IN X 10.0 IN X 5.9 IN
D-1L	4,298 FEET	5,318 FEET	1.3 IN X 14.4 IN X 8.1 IN

CASSETTE STORAGE		
	18 μ m TAPE	13 μ m TAPE
D-1S	1.1X10 ¹¹ BITS	1.5X10 ¹¹ BITS
D-1M	3.4X10 ¹¹ BITS	4.5X10 ¹¹ BITS
D-1L	7.6X10 ¹¹ BITS	1X10 ¹² BITS
RECORD TIME AT 200 MBH/s		
D-1S	8.5 MINUTES	12 MINUTES
D-1M	30 MINUTES	37 MINUTES
D-1L	68 MINUTES	82 MINUTES
TAPE		
COERCIVITY ENHANCED IRON OXIDE 850 Oe		

■ DATATAPE ID-1 DCTR-LP SERIES LABORATORY RECORDER



HARRISS ROBINSON DATATAPE INCORPORATED

24

HIGH PERFORMANCE 19-mm HELICAL SCAN DIGITAL TAPE RECORDERS

■ DATATAPE ID-1 DCTR-A120 SERIES AIRBORNE MIL-E-5400 RECORDERS



TAPE TRANSPORT UNIT
(TTU)



RECORD/REPRODUCE UNIT
(RRU)

HARRISS ROBINSON DATATAPE INCORPORATED

HIGH PERFORMANCE 19-mm HELICAL SCAN DIGITAL TAPE RECORDERS

- **ID-1 CONSERVATIVE DESIGN APPROACH**
 - TRACK SPACING 0.045mm (0.0018")
 - 850 OERSTED GAMMA FERRIC OXIDE TAPE
 - HIGH DATA RATES ACHIEVED TO 50 MBPS
 - MULTIPLE SUPPLIER OF EQUIPMENT DESIGNED TO STANDARDS
- **AMPEX R-90 HELICAL SCAN RECORDER BASED ON USING
AMPEX VPR-300 D-2 ROTARY DIGITAL VIDEO BROADCAST
MACHINE FEATURES**
 - Helical scan recorder
 - 15 Megabyte per second transfer rate
 - IPI-3 interface
 - 300 inch per second shuttle speed
 - 150 inch per second search speed
 - Accepts three sizes of cassettes
 - Small cassette holds 25 gigabytes
 - Medium cassette holds 75 gigabytes
 - Large cassette holds 165 gigabytes
 - Less than one permanent read error per 10¹² bytes read
 - Interleaved Reed-Solomon ECC
 - Read-after-write with selectable rewrite
 - Automatic reread for soft errors

HARRISS ROBINSON DATATAPE INCORPORATED

26

FUTURE DEVELOPMENT

- **HIGH COERCIVITY TAPE**
FROM 850 OERSTED TO: TBD
- **HIGH PERFORMANCE HEADS**
FROM FERRITE TO METAL-IN-GAP, TO METAL LAMINATED HEADS
- **PER CHANNEL DATA RATES FROM:**

50 TO >200 MBPS ∴ 8-CHANNEL MACHINE WILL BE ABLE
TO OPERATE AT OVER 1,600 MBPS (>200 MBPS), A RATE
SUBSTANTIALLY HIGHER THAN REQUIRED FOR HDTV

HARRISS ROBINSON DATATAPE INCORPORATED

27

SUMMARY

- MACHINES REQUIRED FOR STORAGE OF BIT-FILES OF ANY LENGTH ARE MORE DEMANDING THAN STREAMING DISK BACK-UP OR CONTINUOUS VIDEO RECORDING
 - DATA INTEGRITY AND RELIABILITY ARE ABSOLUTE REQUIREMENTS
 - DATA RECORDED ON A CASSETTE IN ONE MACHINE MUST BE REPRODUCIBLE IN ANOTHER MACHINE
- DESIGNS SHOULD BE GENEROUS, e.g., TRACK WIDTH AND BIT DENSITY
- SOUND APPROACH IS TO MINIMIZE FACTORS WHICH CONTRIBUTE TO ERRORS (TAPE QUALITY AND TRACKING) AND INCORPORATE ROBUST ERROR CORRECTION (INTERLEAVING, CODING, AND WRITE-READ AND REWRITE).
- ANSI ID-1/DD-1 MACHINES INCORPORATE THE DESIRABLE FEATURES FOR CURRENT GENERATION HIGH SPEED, HIGH CAPACITY DIGITAL DATA RECORDERS.

HARRISS ROBINSON DATATAPE INCORPORATED

28

N93-13136

**Storage System Software Solutions
For High-End User Needs**

Carole B. Hogan
Director of UniTree Technical Development

Distributed Computing Solutions, General Atomics
San Diego, California

PRECEDING PAGE BLANK NOT FILMED

Storage System Software Solutions For High-End User Needs

Abstract

Today's high-end storage user is one that requires rapid access to a reliable terabyte-capacity storage system running in a distributed environment. This paper discusses conventional storage system software and concludes that this software, designed for other purposes, cannot meet high-end storage requirements. The paper also reviews the philosophy and design of evolving storage system software. It concludes that this new software, designed with high-end requirements in mind, provides the potential for solving not only the storage needs of today but those of the foreseeable future as well.

Introduction

Since the days of computer infancy, people have used computers to generate long-term data, i.e., data that they desired to access over extended periods of time. To store this data and to permit access to it, computers included a system of storage hardware and storage software. Generally, these storage systems provided similar services to all users. The storage hardware consisted of a single storage media type. The storage software provided a convenient interface for users to access and manage data stored on the hardware media.

As the development and use of computers became more specialized and sophisticated, storage needs of users began to differ. Some user needs were simple, for example, large numbers of small files accessed frequently. Other user needs were more demanding, fewer numbers of large files that were accessed with varying frequencies. Even though user storage needs were diverging, storage systems generally continued to offer the same general services, failing to keep pace with differing user requirements.

Today, user storage needs are continuing to diverge and are doing so at a faster rate than in the past. Conventional storage systems, i.e., the systems that have grown up with the first generations of computer users, can no longer even pretend to meet the differing storage needs of current users. However, new storage systems are being provided that can meet the needs of a variety of storage users in today's computing environments and can also adapt to changing needs in the environments of tomorrow.

The remainder of this paper focuses on the particular storage software needs of high-end users, conventional storage systems software and the shortcomings that prevent it from meeting these needs, and the features of the newer storage system software that make it an ideal solution for this user community.

Storage Needs

In general, computer users with high-end needs are those that generate terabytes of long-term data from high-performance processors. Working in a network environment, these users require the data to be placed in a reliable storage system that can be rapidly and conveniently accessed from several points in the network. Because of the finite budgets, the storage system must be able to provide its services by efficient use of continually-advancing storage technology.

While high-end users have several needs from their storage system software, their primary needs are reliability, capacity, performance, hardware flexibility, software flexibility and reasonable cost. Each of these needs is discussed in turn.

Reliability

Reliability is the capability to guarantee that data accepted for storage will be returned in the same state in which it was received, and the capability to provide continuous storage system availability.

The data integrity component of the reliability need, i.e., returning data to the user in the same state in which it was stored, is the traditional hallmark of storage system software, particularly for storage systems designed as long-term archives. If users cannot trust the storage system software to provide data integrity then, regardless of its other attributes, the system is not useful to them.

The system availability aspect of the reliability need is particularly important for real-time applications such as seismic monitoring programs. Storage system software must provide continuous storage system service in order for these applications to successfully store the data received from the real-time source.

Capacity

Capacity is the capability to store increasing amounts of data upon demand.

The capacity requirement is primarily a software requirement to keep pace with developing hardware technology that permits the generation of larger and larger quantities of data. For example, the processor and main memory technology provided in computers from which users generate data is continually improving in speed and size. This allows users to generate more data in shorter periods of time.

Other types of sophistication in data generation techniques also require increasing storage capacity. For example, satellite-based data generation can result in petabyte capacity requirements. Storage system software must be able to meet these capacity requirements upon demand, in order to be useful in these types of computing environments.

The capacity requirement also prevents the placement of limitations on the numbers of files that can be stored or the size of the files that are stored in the system.

Performance

Performance is the capability to provide access to data as quickly as possible.

The need for fast response from storage system software, both on read and write accesses, is especially important when users generate large amounts of data in a network, or distributed, environment. In such an environment, the storage system software usually runs on a processor or processors dedicated to its use while users access the system from a variety of other computers. So in order for the software to provide optimum performance, it must be efficient enough to utilize high-speed networks and high-speed storage devices at their maximum performance capability.

Hardware Flexibility

Hardware flexibility is the capability to adapt to new storage media, and the capability to manage multiple media in the same storage system.

Advances in hardware technology continue to provide faster, cheaper, greater-capacity storage media. In order for users to receive the benefit of these advances in technology, the storage system software must be flexible enough to store, retrieve and manage data on the new media with only minor modifications to the software.

Storage system software must also be able to manage different hardware media types in the same storage system to give users the option of utilizing media with different storage and cost/performance characteristics.

Software Flexibility

Software flexibility is the capability to operate on several software platforms, and the capability to adapt to new application requirements.

Generally, storage system software is designed to interface with the software operating system on a given processor. Because users are often best served by periodic upgrades to different storage processors with different operating systems, the storage system software must be flexible enough to run on a variety of software platforms.

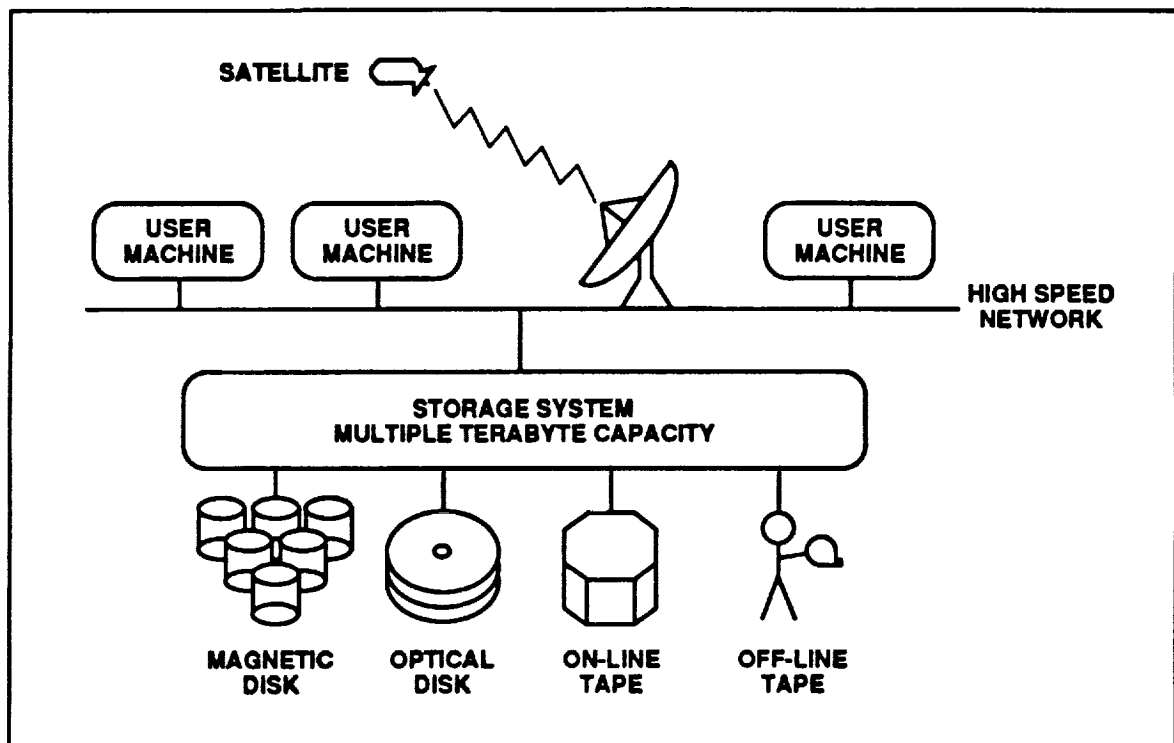
Storage system software must also be able to adapt to new application requirements such as providing the means to identify files by attribute rather than by more traditional means.

Reasonable Cost

Reasonable Cost is the capability to provide the maximum amount of storage within defined budget guidelines.

To be cost effective, the storage systems software must be able to utilize cost-reducing advances in storage hardware technology such as managing cheaper storage media types as they become available. Also, storage system software that can utilize robotically-mounted storage media will permit reduction in the labor costs associated with manually mounted media.

Figure 1 presents an example of a network computing environment that depicts a storage solution for high-end storage needs.



(Figure 1)

Conventional Storage System Software

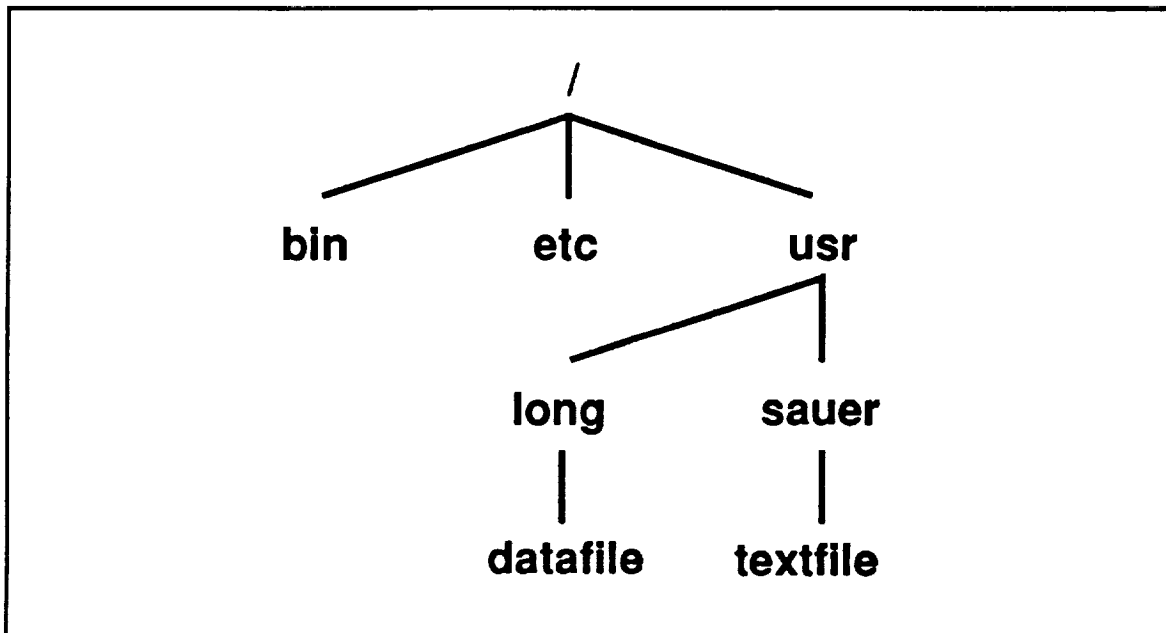
This section of the paper describe conventional storage system software and discusses the shortcomings inherent in this software that prevent it from meeting today's high-end storage needs.

Features

In general, conventional storage system software is designed to manage large amounts of data over long periods of time, and to provide convenient user access to the data. Convenient user access is provided through two abstractions, a file and a human name to identify a file.

The file abstraction allows users to manage their data in convenient units rather than as individual bits. A human name for a file permits users to access the file with that name rather than requiring them to supply the actual address on the storage media where the file is stored. The conventional human name for a file is a directory pathname.

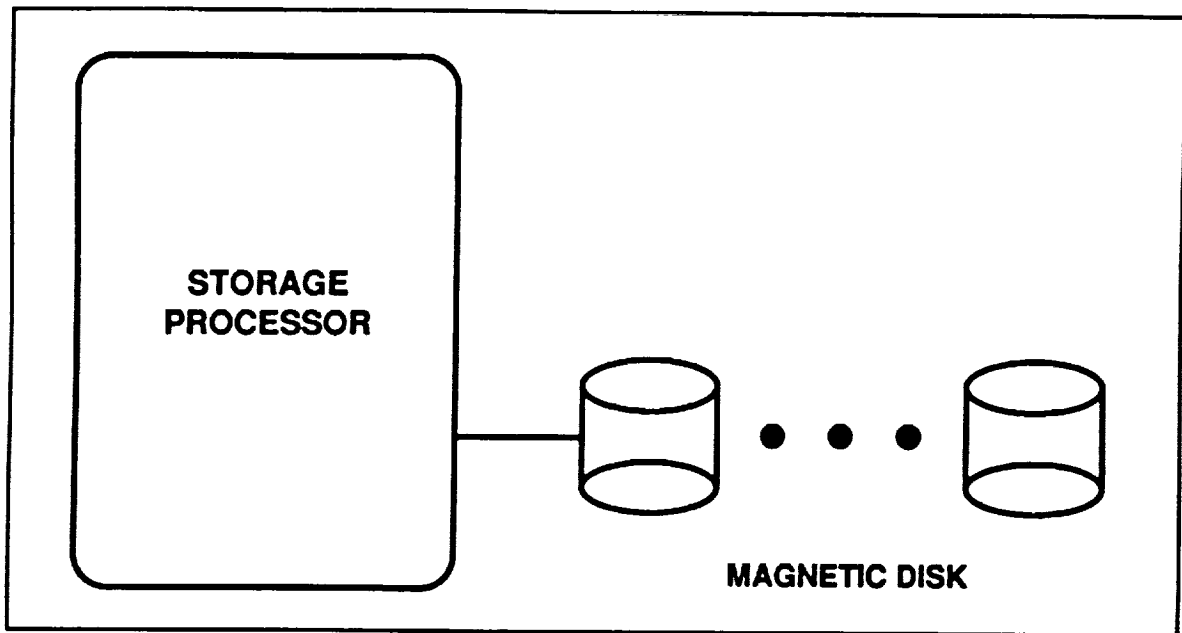
Figure 2 presents a sample directory structure of the UNIX file system, a conventional storage system familiar to many users. The two files in the directory structure shown have the human names `"/usr/long/datafile"` and `"/usr/sauer/textfile"`.



(Figure 2)

Typically, conventional storage system software provide operations on files such as create, open, write, read, and close. It also manages files by storing them on a single storage media type. The conventional storage media is magnetic disk.

Figure 3 shows the hardware configuration for the conventional UNIX file system.



(Figure 3)

Meeting Storage Needs

Generally, conventional storage system software fails to fulfill each of today's high-end storage requirements because it is designed to meet simpler storage needs in simpler computing environments. Some of its specific shortcomings are discussed here.

Reliability

Storage system software provides data integrity by using algorithms that function correctly and by providing mechanisms that preserve data when hardware media fails. While conventional system algorithms function correctly, the software makes no provision for protecting data from media failure, such as providing redundant copies on different media or automatic backup procedures. Generally, assurance of data integrity in these systems is only as good as the reliability of the storage hardware media and the frequency of manual backup of the data to other media.

Further, providing continuous system availability is not a feature of conventional storage system software. For example, the flexibility to manage a single storage database from two copies of the same software running on two different processors is absent. Such a feature would permit continuous service across processor failure.

Capacity

The capacity to meet increasing storage capacity demands is also absent from conventional storage system software due to built-in limitations in the software. First, capacity is limited because the software manages only magnetic disk as the storage hardware media. There is a limit to the amount of disk that can be attached to any processor.

Second, capacity is also limited because conventional software is not designed to continue writing a file to another disk with available space when the first disk becomes full.

This limitation is not only an inconvenience to users and system administrators but it also means that the largest file in the system is restricted to the size of the largest disk.

Third, total storage system capacity is also limited by a software restriction on the number of data blocks that can be addressed in the system. This restriction holds regardless of the number of disks managed by the software. Usually, this addressing restriction limits total storage capacity to hundreds of gigabytes on even the largest computers.

Performance

The greatest cause of performance degradation in conventional storage systems is the necessity for manual storage media management. When magnetic disks become full of data, a human operator must off-load the data to other media, such as tape, in order to create space on the disk for storing more files. Later, when a user desires to access a file that has been moved to tape, the operator must first locate the tape and read the file from the tape back to disk. Once the file is on disk again, the user can access it.

Any software system that requires manual media management will not perform satisfactorily in today's high-end computing environment.

Hardware Flexibility

The hardware flexibility requirement of adapting to new storage hardware media and of managing multiple media types in the same storage system software. Conventional software, as stated above, manages only magnetic disk as the storage hardware media.

Software Flexibility

Conventional storage system software is designed to run on only one type of operating system. Therefore, the same software cannot be used to manage user data if the storage processor is upgraded to one that provides a different operating system. In this case, another brand of storage system software must be used. This usually requires reformatting data from the old system format to the new system format.

Also, conventional storage system software lacks the flexibility to adapt to new application requirements such as new naming conventions for files. Because the conventional mechanism for naming files, i.e., the directory structure, is intertwined with file management, a new naming mechanism, such as a database management system, will require a different brand of storage system software.

Reasonable Cost

Conventional storage system software keeps storage costs relatively high by managing only magnetic disk and by requiring labor-intensive manual storage media management.

The remainder of the paper discusses the philosophy, features and architecture of today's evolving storage system software, and how this software can meet the storage needs of high-end users for the foreseeable future.

Evolving Storage System Software

Recognizing the need for storage system software that is capable of responding to user needs in the computing environments of today and tomorrow, the IEEE Technical Committee on Mass Storage Systems and Technology and the IEEE Storage System Standards Working Group are developing a model for mass storage systems. In its most recent published version, the Mass Storage System Reference Model¹ advocates concepts and an architecture that will allow storage system software to meet user needs for the foreseeable future.

Much of today's evolving storage system software takes its direction from the Mass Storage System Reference Model. Although the Model is not specifically described in this paper, its influence is visible in the philosophy, features and architecture of the evolving storage systems software discussed here.

Philosophy

The philosophy of evolving storage system software states that the software must be flexible enough to meet current application needs and to be easily adapted to changing application needs and hardware technology advances. This philosophy, seemingly unremarkable, deserves special notice because it is the first time storage software has been dedicated to meeting current and changing needs and technology. Certainly, conventional storage system software is designed without this motivation in mind and, as a result, fails to satisfy high-end user needs.

Key Features

There are several key features of evolving storage system software. These features are described here. Taken together, they provide a flexible storage system.

Redundancy

Redundancy is the ability to continue storage service in the event of hardware failure such as processor or storage media failure. If a storage processor fails, access to the storage system is threatened. If hardware media fails, access to data on that media is threatened.

Evolving storage system software is designed to allow continuance of service in the event of processor failure by providing the capability to access the same storage database from a standby processor. If the primary processor on which the storage system is operating fails, the standby processor can be started and the storage system run from that hardware platform.

Evolving storage system software also provide the capability to store redundant copies of data on different devices of the same media such as different tapes. If this feature is taken advantage of and a tape fails, data is not lost because copies exist on other tapes.

No Software Limits

A lack of software limits means that the capacity of the storage system is limited not by the software, but only by the amount of storage hardware media allocated to it.

Evolving storage system software places no limitation on the length of files, the number of files, the lengths of human names for files or the number or size of directories, it that is the naming mechanism, that the storage system can contain. Unrestricted file length is particularly important since it means that a single file can span devices of the same storage media type such as disk or tape, and is no longer limited to the capacity of a single disk or tape.

Storage Media Hierarchy

Evolving storage system software has the capability to manage a storage media hierarchy. A hierarchy is different types of storage media configured in a layered or hierarchical fashion and managed as part of a single storage system. In general, the media in a storage hierarchy are layered according to speed of access, with the faster and generally more expensive media in the higher layers and the slower and generally cheaper media in the lower layers of the hierarchy.

For example, a storage hierarchy might consist of magnetic disk, optical disk, on-line magnetic tape and off-line magnetic tape. Data is moved between these layers according to site-configurable parameters and user access.

Automatic Migration and Caching

Evolving storage system software offers the features of migration and caching to automatically move data between the layers of the storage media hierarchy. Migration is the movement of data downward in the layers of the hierarchy. Caching is the movement of data upward in the layers of the hierarchy. The software moves the data automatically, without human operator intervention, according to site-configurable parameters and user access.

Migration parameters determine how often files are migrated to lower layers and which files are migrated. Caching upward in the hierarchy is triggered by user access of files that have migrated to lower levels. Because of migration and caching, active files, i.e., those currently being accessed, reside on the higher, faster access layers of the hierarchy, while inactive files reside on the lower, slower media.

Distributability

Evolving storage system software consists of modules that are capable of being distributed among several processor in a network environment. This means that the storage system software is not confined to a single processor but can run on several processors simultaneously as part of a single storage system. Distributability permits load balancing among components of the system and faster access to data.

Separation of Control and Data

User access to a storage system consists of commands transmitting data (e.g., write commands) and commands receiving data (e.g., read commands). Each command consists of two parts, control and data. The control component is the action to be taken, e.g., write or read. The data component is the actual data to be written or read.

Evolving storage system software separates the control and data components of storage access commands into two distinct entities that can be transmitted separately to and from the storage system. This separation permits control and data to flow over separate network paths.

For example, control can flow over slower network paths while data can flow over faster network paths. This type of separation permits faster access to the storage system. Since data transmissions are generally much larger than control transmissions, transmitting data over a separate path or paths can result in faster performance in general.

Extensibility

Evolving storage system software is extensible. Extensibility is the capability to easily integrate additional types of storage media into the storage media hierarchy. Easy integration of additional media permits sites to take advantage of new storage hardware technology as it develops.

For example, a site may begin with a storage media hierarchy consisting of magnetic disk, on-line tape and off-line tape. Because of the extensibility of the storage software, optical disk can be integrated into the hierarchy, for example, between the magnetic disk and on-line tape layers, when the site decides to take advantage of the features offered by optical disk.

Separate Naming Mechanism

Evolving storage system software is designed to easily adapt to new naming conventions so that users can access their data in the manner that is most efficient for their

applications. For example, if applications at a given site generate many thousands of files, naming files by directory pathname can become inefficient. Instead, naming files by attribute such as time of creation may be preferred.

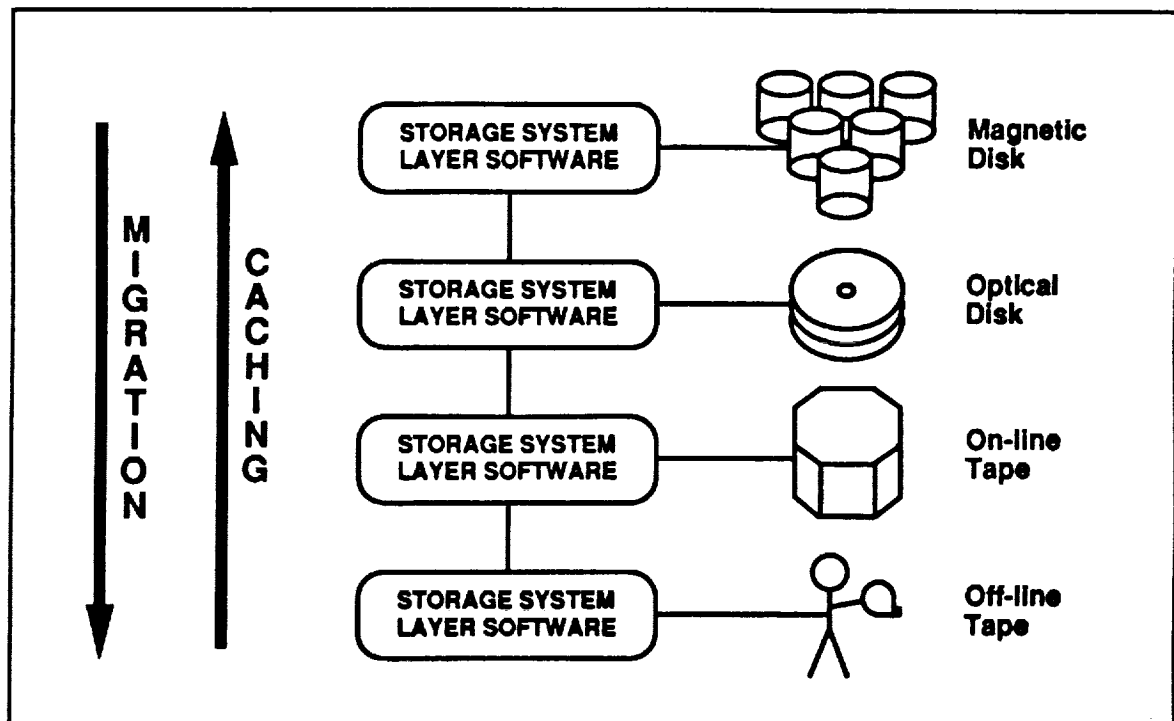
Evolving storage system software will permit this transition in naming mechanism from directory pathname to file attribute because the software modules that map names to files are separate from, and peripheral to, the modules that manage the files. Therefore, to transition to a new naming mechanism means only to change a few peripheral modules rather than the entire storage system software.

Portability

Evolving storage system software is portable. It is designed to run on a variety of software operating systems. This means that if the storage processor is upgraded to a different manufacturer with a different operating system. It also means that the same storage software can manage a single distributed storage system by running on a variety of different processors with different operating systems.

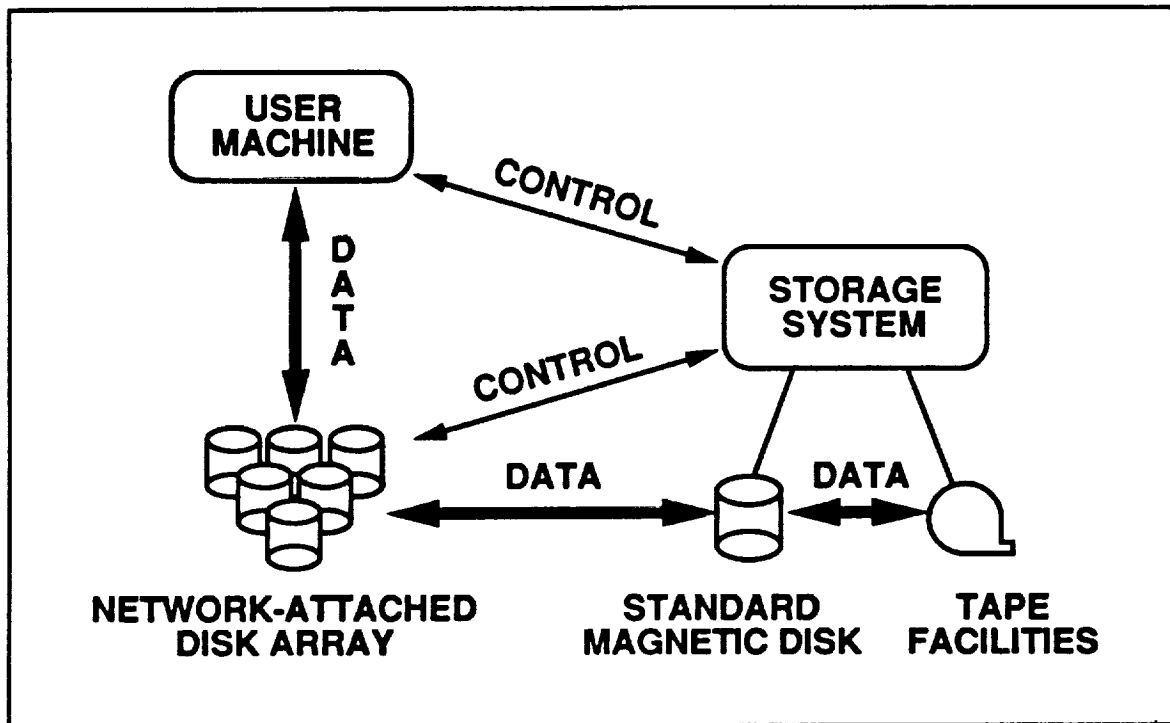
Architecture

The architecture of evolving storage system software is presented here graphically. Figure 4 shows the software and hardware layers in a sample storage media hierarchy. Each layer of media is managed by a different set of software modules.



(Figure 4)

Figure 5 shows the separate network paths that control and data can take between the user and the storage system. In this Figure, data is stored in a three-layer hierarchy, consisting of a network-attached disk array, standard magnetic disk and tape. The last two layers in the hierarchy are attached to the storage processor. Storage system software is running on the disk array controller as well as the storage processor itself. In this way, data can flow directly between the user machine and the disk array without passing through the storage machine.



(Figure 5)

Meeting Storage Needs

The features and architecture of evolving storage system software permit it to meet the storage needs of high-end computer users. The manner in which the software meets these needs is discussed here.

Reliability

System availability and data integrity are provided by the redundancy features of evolving storage system software. Redundancy provided the capability to continue service from a standby processor. It also preserves data in the event of damage or loss of a storage media device such as a tape by providing the capability of storing multiple copies of the same data on different tapes.

Capacity

Evolving storage system software provides the potential for unlimited storage capacity by managing a hierarchy of storage media and by a lack of limits in the number of files that can be stored, the size of files that can be stored, and so forth. Multi-media management in a hierarchical fashion allows tape to be included in the storage system. Tape as a storage medium, coupled with a lack of limitation in file size and numbers, means that the system capacity can grow to as many tapes as a site purchases.

Performance

Evolving storage system software provides performance through the features of automatic migration and caching, separation of control and data and distributability. Automatic migration and caching eliminate the performance degradation caused by conventional manual media management. Separation of control and data permit data to flow to and from the storage system over faster network paths, improving performance in general. Distributability allows components or layers of the storage software to run on different processors, permitting load balancing and faster access to the system as a whole.

Hardware Flexibility

The extensibility feature allows evolving storage system software to provide hardware flexibility. With this flexibility, new storage media can be integrated into the storage hierarchy. The ability to integrate new media permits users to take advantage of advancing hardware technology.

Software Flexibility

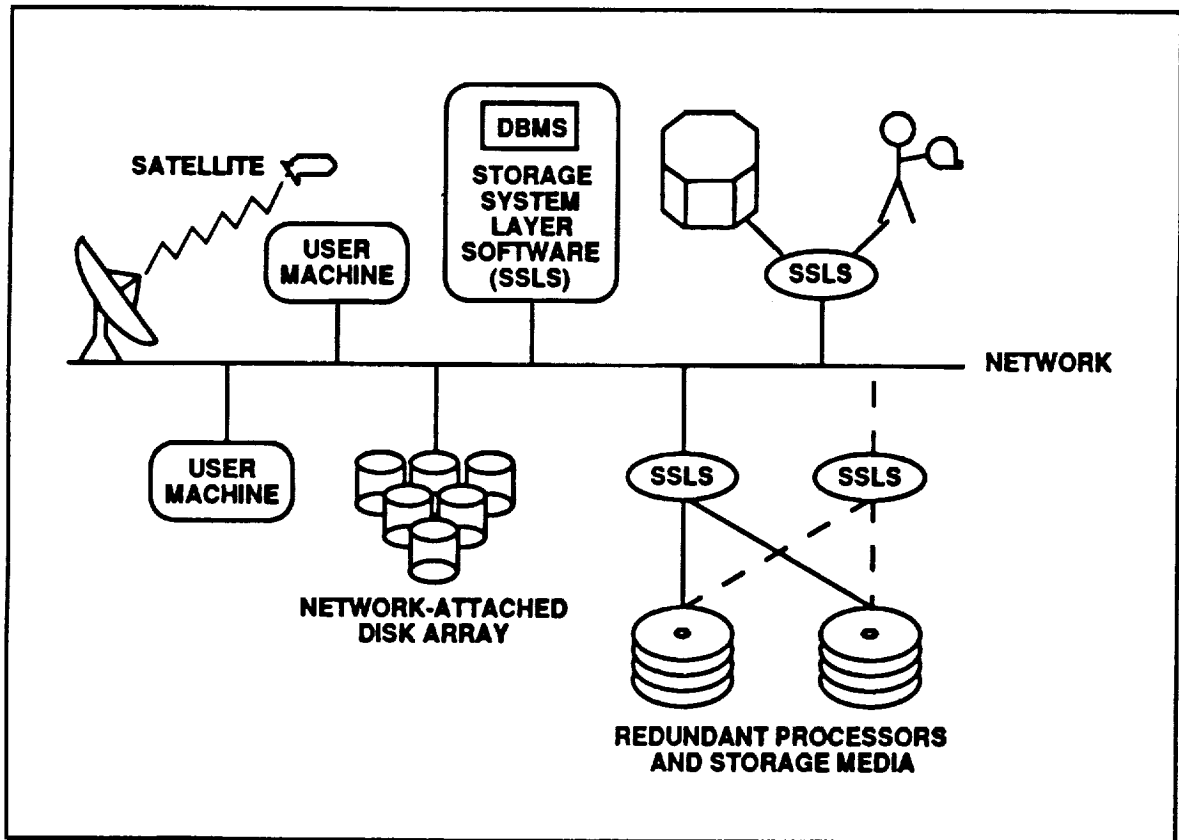
Software Flexibility is provided by several features of evolving storage system software, including a lack of software limitations, separate modules that map human file names to the files themselves, distributability and portability. Portability is the flexibility to run on a variety of software operating systems.

Reasonable Cost

Reasonable cost is provided by evolving storage system software through the features of storage media hierarchy management, automatic migration and caching, extensibility and portability. Managing data through a storage media hierarchy means that less money need be spent on expensive storage media because only active files will be stored on this higher layers in the hierarchy. Generally, active files are a small subset of all files in a storage system. Automatic migration and caching eliminates the cost associated with manual storage media management.

Extensibility allows cheaper storage media to be integrated into the system as it becomes available. Portability accommodates less expensive open system solutions to storage needs since the same storage software can run on a variety of software platforms that can be combined into a single storage system.

Figure 6 shows an example of an evolving storage system managing storage needs in a high-end, distributed computing environment. In this example, the naming mechanism is a database management system. The layers of the storage hierarchy consist of a disk array, optical disk and on-line and off-line tape. These layers are distributed among different processors in the network. Redundancy is provided in the optical disk layer of the storage hierarchy by a standby processor (dotted line) and redundant sets of optical disk.



(Figure 6)

Conclusion

Users in computing environments with high-end storage requirements must turn away from using conventional storage system software if these requirements are to be met. Conventional software, designed for simpler computing environments with simpler storage requirements, cannot satisfy today's high-end storage needs.

Evolving storage system software, influenced by the IEEE Mass Storage System Reference Model, is designed to meet high-end storage requirements for the foreseeable future. It is in these systems that today's high-end users will find solutions to their storage software needs.

Acknowledgement

The contributions of Michael Hardy, Products Manager, Distributed Computing Solutions, to this paper are gratefully acknowledged.

References

1. Sam Coleman and Steve Miller, "Mass Storage System Reference Model: Version 4 (May, 1990)," Tenth IEEE Symposium On Mass Storage Systems, Monterey, California, May, 1990.

Storage System Solutions
NSSDC Conference on Mass Storage Systems

Carole B. Hogan

Director of UniTree Technical Development
Distributed Computing Solutions (DISCOS)

July 23, 1991



Overview:

- **Storage system definition**
- **Storage needs**
- **Conventional storage systems**
- **Evolving storage systems**

PRECEDING PAGE BLANK NOT FILMED



Storage System Definition:

- **Facility that**
 - **manages**
large amounts of data
over long periods of time
 - **provides**
convenient use of data



Storage Needs:

- **Reliable storage**
- **Large amounts of data**
- **Rapid response**
- **Hardware flexibility**
- **Software flexibility**
- **Reasonable cost**



Reliable Storage:

- **Capability to provide system availability and data integrity**
 - **real-time applications**
 - **long-term archive**



Large Amounts of Data:

- **Capacity to store and provide access to increasing amounts of data**
 - **satellite-based data acquisition**
 - **faster processors**



STORAGE SYSTEM SOLUTIONS

Rapid Response:

- **Capability to provide access to data as quickly as possible**
 - **high-speed networks**
 - **high-speed storage devices**



STORAGE SYSTEM SOLUTIONS

Hardware Flexibility:

- **Capability to adapt to new storage media**
 - **optical-based**
 - **helical scan tape**



STORAGE SYSTEM SOLUTIONS

Software Flexibility:

- **Capability to operate on several software platforms**
- **Capability to adapt to new application requirements**
 - **database management system access**



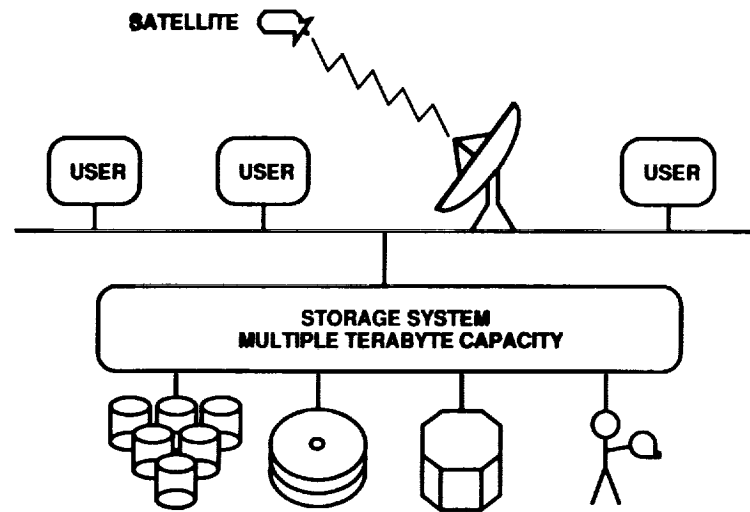
STORAGE SYSTEM SOLUTIONS

Reasonable Cost:

- **Provide maximum amount of storage within budget guidelines**
 - **automated storage media management**



STORAGE SYSTEM SOLUTIONS



STORAGE SYSTEM SOLUTIONS

Conventional Storage Systems:

- **Description**
- **Example**
- **Shortcomings**



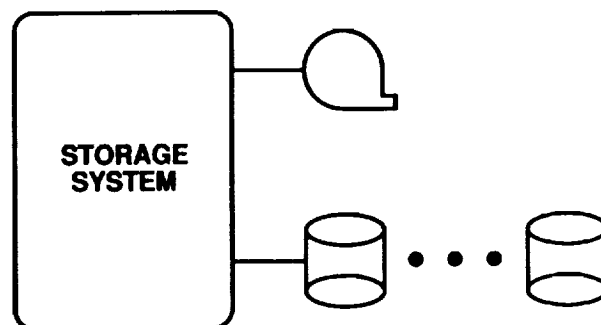
Conventional Storage System Description:

- **Provides**
 - **long term storage and retrieval of user data**
 - **file abstraction for users to manage data**
 - **directory pathname as the mechanism to name files**
 - **open, write, read, and close operations on files**
 - **management of magnetic disk storage media for files**



Conventional Storage System Example:

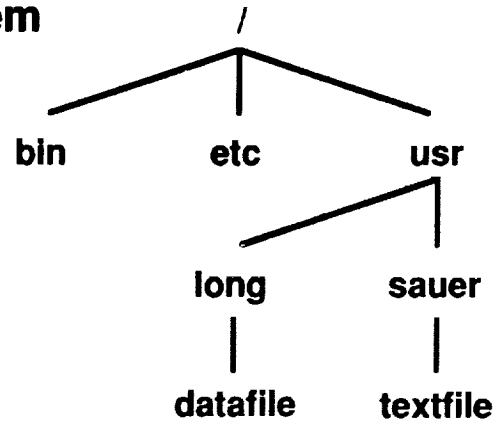
- **UNIX File System**





Conventional Storage System Example:

- **UNIX File System**



Reliability:

- **Only as good as**
 - **reliability of magnetic disk**
 - **frequency of backup**



Conventional Storage System Shortcomings:

- **Reliability**
- **Capacity**
- **Performance**
- **Hardware flexibility**
- **Software flexibility**
- **Cost**



Capacity:

- **Manages single storage media type**
 - **few per machine**
- **Provides no mechanism to continue writing to another disk which has available space**
 - **limited file size**
- **Uses software feature that limits total capacity**
 - **100s of gigabytes**



STORAGE SYSTEM SOLUTIONS

Performance:

- **Manual storage media management**
 - **operator off-loading data from disk to tape and back again**



STORAGE SYSTEM SOLUTIONS

Hardware Flexibility:

- **Magnetic disk as the only storage media type**



STORAGE SYSTEM SOLUTIONS

Software Flexibility:

- **No capability to adapt to new application requirements such as new naming conventions**



STORAGE SYSTEM SOLUTIONS

Cost:

- **More and more expensive disk**
- **Labor-intensive storage media management**



Evolving Storage Systems:

- **Philosophy**
- **Key features**
- **Architecture**
- **Meeting storage needs**
- **Example**



Philosophy:

- **Flexible storage system that meets current application needs and is easily adapted to changing application needs and hardware technology advances**



Key Features:

- **IEEE Mass Storage System Reference Model**
- **Redundancy**
- **No software limits**
- **Storage media hierarchy**
- **Automatic migration and caching**
- **Distributability**
- **Separation of control and data**
- **Extensibility**
- **Separate naming mechanism**
- **Portability**



Redundancy:

- **The ability to continue service in the event of hardware failure such as processor or storage media failure**



No Software Limits:

- **Unlimited length files**
- **Unlimited number of files**
- **Unlimited length file names**
- **Unlimited number of directories**



Storage Media Hierarchy:

- **Different types of storage media configured in a layered or hierarchical fashion and managed as a single storage system**



Automatic Migration and Caching:

- **Movement of data between the media layers of the storage hierarchy according to administrative policy**
- **Migration is data movement downward in the hierarchy**
- **Caching is data movement upward in the hierarchy**



Distributability:

- **Storage system components running on different processors within a network**



Separation of Control and Data:

- **The ability to transmit control and data over separate network paths**



Extensibility:

- **The ability to easily integrate additional types of storage media into the storage hierarchy**



Separate Naming Mechanism:

- **The ability to easily adapt to new naming conventions**



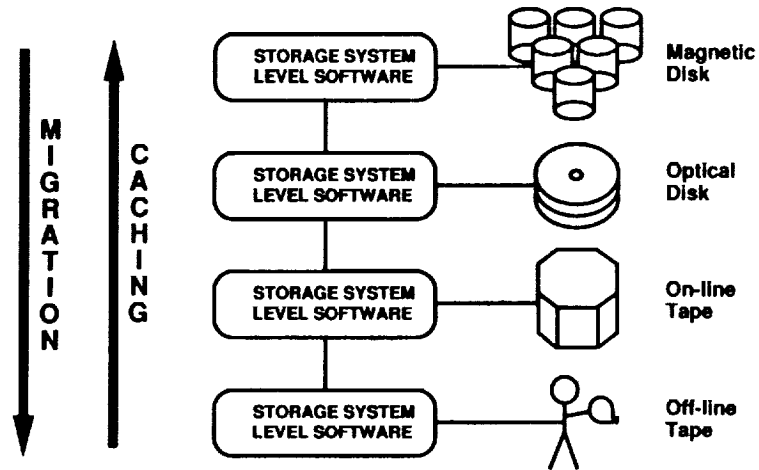
Portability:

- **The ability to run on a variety of software platforms**



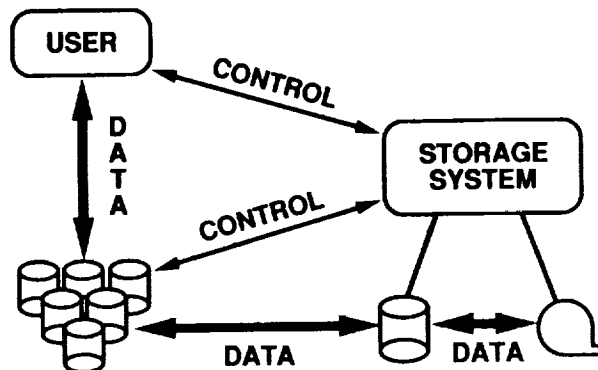
STORAGE SYSTEM SOLUTIONS

Architecture:



STORAGE SYSTEM SOLUTIONS

Architecture:





Meeting Storage Needs:

- **Reliability**
- **Capacity**
- **Performance**
- **Hardware flexibility**
- **Software flexibility**
- **Reasonable cost**



- **Reliability:**
 - **Redundancy**
- **Capacity:**
 - **Storage media hierarchy**
 - **No software limits**



STORAGE SYSTEM SOLUTIONS

- **Performance:**
 - **Automatic migration and caching**
 - **Separation of control and data**
 - **Distributability**
- **Hardware Flexibility:**
 - **Extensibility**

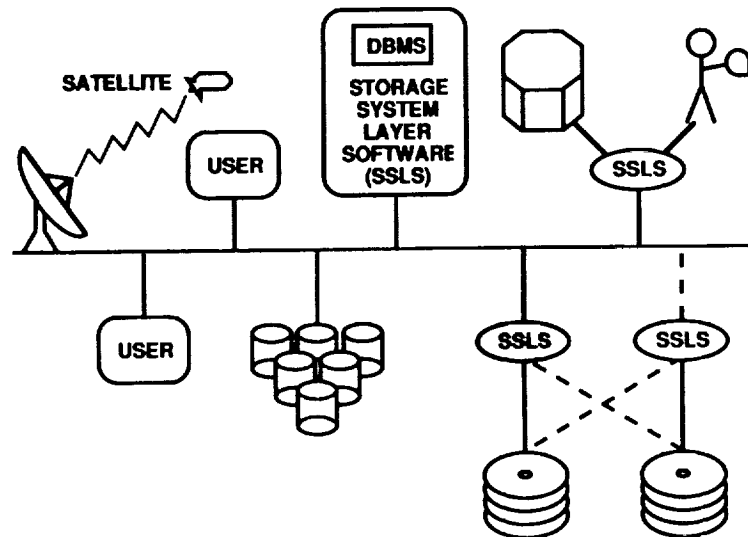


STORAGE SYSTEM SOLUTIONS

- **Software Flexibility:**
 - **No software limits**
 - **Separate naming mechanism**
 - **Distributability**
 - **Portability**
- **Reasonable Cost:**
 - **Storage media hierarchy**
 - **Automatic migration and caching**
 - **Extensibility**
 - **Portability**



STORAGE SYSTEM SOLUTIONS



File Servers, Networking, and Supercomputers

Reagan W. Moore

San Diego Supercomputer Center
San Diego, California

Abstract

One of the major tasks of a supercomputer center is managing the massive amount of data generated by application codes. A data flow analysis of the San Diego Supercomputer Center is presented that illustrates the hierarchical data buffering/caching capacity requirements and the associated I/O throughput requirements needed to sustain file service and archival storage. Usage paradigms are examined for both tightly-coupled and loosely-coupled file servers linked to the supercomputer by high-speed networks.

Introduction

The file server capacity requirements are most strongly driven by the CPU power of the central computing engine. The workload that can be sustained on the supercomputer is ultimately limited by the ability to handle the resulting I/O. At the San Diego Supercomputer Center, the central computing resource is a CRAY Y-MP8/864 supercomputer with a peak execution rate of 2.67 Gflops, capable of generating up to

$$2.67\text{E}+9 \text{ operations/sec} * 8 \text{ Bytes/operation} * 86,400 \text{ sec/day}$$

or 1.8 Petabytes per day. In practice, the actual data generation rate is determined by the workload characteristics. The two major sources of I/O are application disk I/O and job swapping to support interactive use. At SDSC, the batch load averages 8 GBytes of executable job images, while the interactive load peaks at 120 simultaneous users. The batch load is sufficiently large that the idle time on the supercomputer has averaged 1.5% of the wall clock time over the last 6 months. While maximizing CPU utilization has been an explicit goal at SDSC, this has also increased the total amount of data that must be manipulated.

Data Flow Analysis

Not all generated data are archived and not all archived data are saved forever. A data flow analysis is necessary to understand the characteristics of the I/O, including the amount of data actually generated, the length of time over which the data are accessed, and the rate at which the data are moved through multiple caching levels. A simple analysis of the data flow can be used to illustrate the results of changing data access methods, increasing processing power, or improving network bandwidth. In particular, the data flow patterns are expected to be different for loosely-coupled user-initiated file archiving than for automated file servers tightly-coupled to the supercomputer CPU.

The current archival storage system in use at SDSC is DataTree which supports user-initiated file archiving. This system acts as the archival storage file server for the CRAY supercomputer and is accessed through a 100 Mbits/sec FDDI ring. Data generated on the CRAY Y-MP8/864 are ultimately stored on 3480 cartridge shelf tape. There are five levels of I/O buffering or caching, including a 1 GByte Solid State Storage Device, 42 GBytes of CRAY disk local to the supercomputer, 70 GBytes of archive disks, a 1.2 TByte tape robot, and 2 TBytes of manually mounted shelf tape. Table 1 illustrates the caching hierarchy. As expected, the amount of data moved towards the lowest archival storage level decreases as the required storage life of the data increases. Data resides on the SSD for periods on the order of minutes, on CRAY disk for up to two days, on archival storage disk for up to several weeks, in the tape robot for several months, and finally on shelf tape for years. The amount of data moved per day between each level varies from 1.5 TBytes/day through the SSD to CRAY disk, 14 GBytes/day through archival storage, 9 GBytes/day through the tape robot, and 2 GBytes/day to shelf tape. The residency time at any level may be estimated by dividing the size of the cache by the input I/O rate to the cache. This closely matches measured data residency times.

Local CRAY Disk Cache

1-146

Archival Storage Disk Cache

The true long-term data generation rate is governed by how fast data are migrated to archival storage. On the DataTree archival storage system in use at SDSC, archiving of files is a user initiated process. Users explicitly choose which files to archive or retrieve. Typically 14 GBytes/day of data are transferred between the CRAY disks and the archival storage system of which one third is data written to storage. This amount of data flow is only 1/7 of that needed to migrate the data that survive on CRAY disk to archival storage. Thus about 1.4% of the total amount of data written to CRAY disk is archived. The archival storage disks form an effective cache between long-term storage on cartridge tape within the tape robot and the CRAY local disks. The hit ratio for archival storage data being retrieved from the archival storage disks is typically 92%.

Archival Storage Tape Caches

The average data transfer rate needed to support archival storage is 0.16 MBytes/sec. This should be compared with the observed sustainable archival storage data rates of 0.6 MBytes/sec supported by DataTree running on an Amdahl 5860 across 4.5 MBytes/sec I/O channels connected to a 12.5 MBytes/sec FDDI backbone network. During periods of heavy usage, the average transfer rate does approach the peak rate.

Long term archival storage to tape occurs both directly from the CRAY disk for large files (sizes greater than 200 MBytes) and by automatic data migration from the archival storage disks. The tape robot serves mainly as a data cache. Data currently reside about 15 months before migrating to shelf cartridges. Data caching attributes can be tracked by the fraction of tape mounts done manually. Typically the 1.2 TByte tape robot processes 85-90% of the tape mounts. The rate at which data are migrated from the tape robot to shelf tape is roughly 2/3 of the rate at which data are written to the robot. This ratio may approach one as data in the robot mature.

This data flow analysis demonstrates some interesting attributes of loosely-coupled user-initiated archival file storage systems.

10% of the generated data is stored temporarily on CRAY local disk, 1.4% of the generated data is written to archival storage, and 0.6% of the generated data is eventually transferred to long term shelf tape. Given the need to explicitly save files, users selectively store a fraction of their output.

The multiple levels of the storage hierarchy serve mainly as caches with more data flowing into a given cache than flows out to lower caching levels.

The amount of data read at each caching level is substantially higher than the amount written with the ratio varying from 4:1 for the highest speed cache on the SSD down to 2:1 for archival tape storage.

The above data flow analysis is typical only of user-initiated archival storage. If an automated archival storage scheme is used for supporting the CRAY disks, the amount of data that are archived could grow substantially. This can seriously impact the ability to adequately handle the I/O if the archival storage hardware environment is operating with relatively small safety margins. Pertinent safety factors are:

cache residency time versus the latency that a data buffer is amortizing,

cache residency time of data files on local CRAY disks versus the time needed for the application to complete, and

sustainable I/O rate versus the peak I/O demand rate.

If any of these factors drop below one, the system will become severely congested and may even fail. At SDSC, all of these safety margins are relatively small. Due to the limited amount of CRAY local disk space, the residency time of files on CRAY disk is comparable to the wall clock time needed to complete an application run for large codes. The weekly average required I/O rate to access files on the archival storage system is 1/4 of the peak observed sustainable rate. Hourly averages of the required I/O rate approach the peak sustainable rate. A usage paradigm shift that increases the I/O load could seriously stress the archival storage system at SDSC.

File Server Paradigm Shifts

Three possible usage paradigm shifts are being investigated at SDSC, two of which are related to file servers tightly coupled to the supercomputer CPU power. The first is a research project funded by the National Science Foundation and DARPA through the Corporation for National Research Initiatives. Prototypes of tightly coupled applications distributed across supercomputers connected by a gigabit/sec network are being developed, including the linkage of an application to the equivalent of a database interface to archived data. The second is a project to investigate the feasibility of incorporating the local CRAY disk and the SSD as caches directly controlled by the archival storage system. The third is the modeling of the impact on the archival storage system of an upgrade to a 100 Gigaflop/sec supercomputer.

High-speed Remote Access

The CASA Testbed is a collaborative effort between the California Institute of Technology, the Jet Propulsion Laboratory, the Los Alamos National Laboratory, and the San Diego Supercomputer Center. One objective is to demonstrate a distributed application efficiently utilizing two supercomputers while simultaneously using a substantial fraction of the gigabit/sec wide area network linking the computers. Simultaneously maximizing bandwidth utilization and CPU utilization requires minimizing the protocol overhead used for the data transmission[1]. The effective bandwidth for the optimal application is given by

$$B / (1 + O * B)$$

where B is the peak bandwidth (bits/sec) and O is the network protocol overhead measured in seconds of overhead per bit transmitted. For high speed networks, network protocol overhead becomes a critical limiting parameter. For present CRAY supercomputers, the network protocol overhead can require the execution power of an entire CPU to support TCP/IP at 700 Mbits/sec.

Given that a suitable file transport protocol is devised with a small enough protocol overhead, the issue of latency across wide area networks may be the next limiting factor. Since the speed of light is finite, data access delays between SDSC and LANL are as great as disk seek times. Efficient access of remote file systems must then cope with buffering data in addition to caching data. The amount of data shipped between an application and a remote database interface to archival storage must be large enough to amortize the data access delay. Depending on the protocol, the amount of data sent may need to be as large as

$$2 L * B$$

where L is the round trip latency measured in seconds. For a LANL/SDSC application running at 800 Mbits/sec, this is still feasible, requiring buffering on the order of 8 MBytes.

Integrated Local and Archival File Systems

Integrating the local file system into the archival storage file system will substantially increase the amount of data that must be processed by the archival storage software. As seen in the SDSC data flow analysis, the amount of data transferred between the supercomputer and the local disks is more than a factor of 1000 larger than the amount transferred to archival storage. Efficiently handling this increase in data rates will require differentiating between "reliable" local file transport and "unreliable" transport across a local network. By scaling the network protocol overhead needed to support TCP/IP at 700 Mbits/sec by the average CRAY local disk bandwidth derived in the data flow analysis, an estimate can be made of the protocol overhead increase. With no protocol enhancements, an additional 20% of a single CPU would be needed to support the archival and local file system integration. This indicates the need for the integrated system to recognize heterogeneous network environments.

An additional complication is that if all of the generated data stored temporarily on CRAY disk is automatically archived, the data flow from local CRAY disk to archival storage could increase by up to a factor of seven. Files written to the scratch /usr/tmp file system require different backup than files written to permanent home directories. An integrated local file system and archival storage file system must allow for a non-uniform usage pattern.

CPU Execution Rate Dependence

A possible ameliorating effect is that as supercomputers become faster, it may become more cost effective to recompute rather than save data. A supercomputer with a sustained execution rate of 100 Gigafllops is expected to be available by 1995. Assuming the data storage patterns remain the same, the I/O generated by such a machine can be estimated by scaling the results of the data flow analysis by the increase in the execution speed, which is roughly a factor of 3000. The cache sizes and I/O communication rates then become:

SSD	3 TBytes	50 GBytes/sec
Local disk	126 TBytes	50 GBytes/sec
Archive disk	210 TBytes	450 MBytes/sec
Shelf tape	6000 TBytes	60 MBytes/sec

The archival storage communication rates need to be decreased by a factor of 10 to become technically feasible. Thus a paradigm shift towards the dynamic regeneration of simulation output may become inevitable.

Acknowledgement

This work was funded in part by the National Science Foundation under Cooperative Agreement Number ASC-8414524 and Grant Number ASC-9020416.

References

1. Moore, Reagan, "Distributing Applications Across Wide Area Networks," General Atomics report GA-A20074, April 1990.

Table 1
Hierarchical Data Caching Levels

Caching Level	I/O per Day	Data Rate	Capacity	Utilization	Residency Period
SSD	1.5 TB	17 MB/s	1 GB	85-100%	minutes
CRAY Disk	1.5 TB	17 MB/s	42 GB	85-90%	days
Archive Disk	5 GB	0.05 MB/s	70 GB	98%	weeks
Tape Robot	9 GB	0.10 MB/s	1.2 TB	68%	months
Shelf Tape	2 GB	0.02 MB/s	2 TB	70%	years

File Servers, Networking, and Supercomputers

Reagan W. Moore

**San Diego Supercomputer Center
San Diego, California**



SAN DIEGO SUPERCOMPUTER CENTER

Archival Storage Systems as File Servers

- **Examine Hierarchical Caching systems**
 - Capacity requirements
 - I/O requirements
- **Based on Usage at SDSC**
 - Archiving supercomputer generated data



SAN DIEGO SUPERCOMPUTER CENTER

File System Usage Paradigms

- **Loosely-coupled to CPU**
 - User initiated file transfers to archival storage
- **Tightly-coupled to CPU**
 - NFS access
 - Integrated local and archival file systems



SAN DIEGO SUPERCOMPUTER CENTER

SDSC Archival Storage Environment

- **Data Generated by CRAY Y-MP8/864 Supercomputer**
- **FDDI 100 Mbits/sec backbone**
- **DataTree Archival Storage System on an Amdahl 5860**



SAN DIEGO SUPERCOMPUTER CENTER

Five Levels of Data Caching

- **Solid State Storage Device (SSD)**
 - 1 GB, 1.2 GB/s access from memory
- **CRAY local disk**
 - 42 GB, 10 MB/s access per disk
- **Archive storage disk**
 - 70 GB, 0.6 MB/s access across FDDI
- **STK tape robot**
 - 1.2 TB, 0.6 MB/s access across FDDI
- **Shelf cartridge tape**
 - 2 TB, 0.6 MB/s access across FDDI



SAN DIEGO SUPERCOMPUTER CENTER

SDSC Workload Characteristics

- **Application Disk I/O**
 - Generated by an average batch load of 8 GBs of executable jobs
- **Job Swapping**
 - Generated by up to 120 interactive users
- **User-initiated File Archiving**
 - Partial archiving of supercomputer data



SAN DIEGO SUPERCOMPUTER CENTER

Data Flow Analysis

- **Track Data Through the Multiple Caches**
 - Cache utilization
 - Hit rate
 - I/O throughput
 - Fraction of peak rate
 - File residency time
- **Identify Caching versus Data Buffering**



SAN DIEGO SUPERCOMPUTER CENTER

SDSC Data Flow

Cache Level	Capacity (GB)	Utilization
SSD	1	85%
CRAY disk	42	90%
Archive disk	70	98%
Tape robot	1200	68%
Shelf tape	2000	70%



SAN DIEGO SUPERCOMPUTER CENTER

SDSC Data Flow

Cache Level	Residency Time	Fraction saved of total I/O written from SSD
SSD	(seconds)	100%
CRAY disk	30 hours	10%
Archive disk	4 weeks	1.4%
Tape robot	15 months	1.4%
Shelf tape	5 years	0.6%



SAN DIEGO SUPERCOMPUTER CENTER

SDSC Data Flow

Cache Level	I/O per Day (GBytes)	Data Rate (MBytes/sec)
SSD	1500	17
CRAY disk	1500	17
Archive disk	5	0.05
Tape robot	9	0.10
Shelf tape	2	0.02



SAN DIEGO SUPERCOMPUTER CENTER

Data Caching Versus Data Buffering

- **SSD Cache Used for Both**
 - /root file system and Interactive swap space are cached
 - Hit rate for accesses is 99%
 - /usr/tmp file system is buffered
 - Hit rate for accesses is 75-85%



SAN DIEGO SUPERCOMPUTER CENTER

File Server Safety Factors

- **Cache Residency Time versus Latency Amortization Time**
- **Cache Residency Time versus File Usage Time**
- **Sustainable I/O Rate versus Peak I/O Demand Rate**



SAN DIEGO SUPERCOMPUTER CENTER

File Server Paradigm Shifts

- **Changes in Funtionality May Require Usage Paradigm Shift**
 - High-speed remote access
 - Integration of local and archival file systems
 - Very high-speed supercomputers



SAN DIEGO SUPERCOMPUTER CENTER

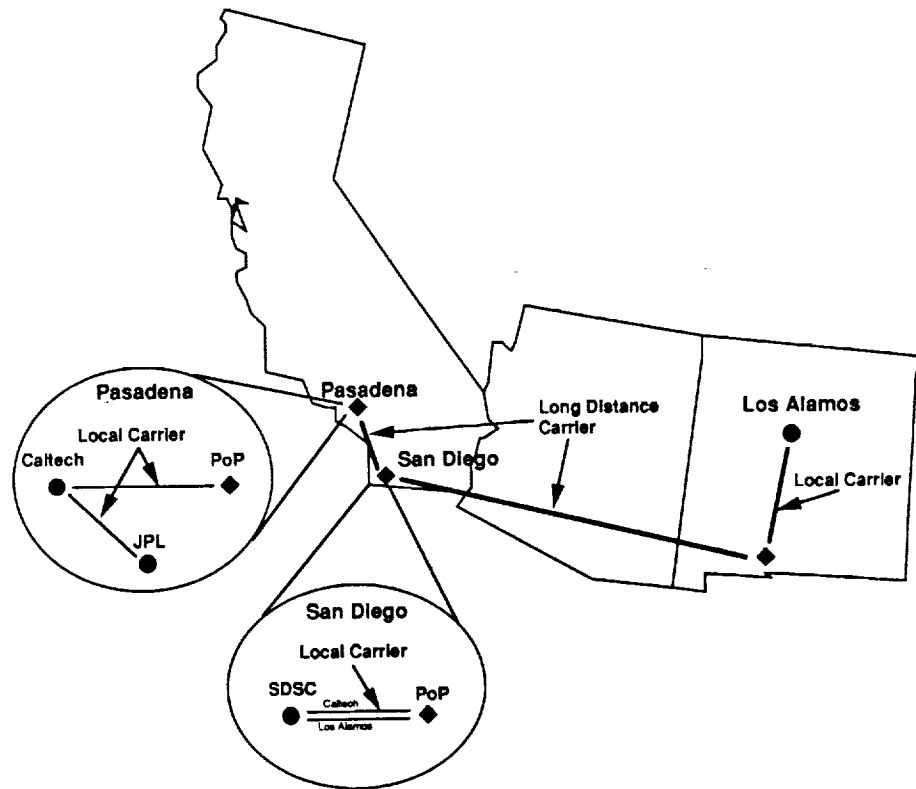
CASA Gigabit/sec Testbed

- **Collaboration between CalTech, JPL, LANL, SDSC**
- **Demonstrate Tightly Coupled Distributed Applications Linked by Gigabit/sec Wide Area Network**
 - Remote access of archived data through database interface



SAN DIEGO SUPERCOMPUTER CENTER

CASA GIGABIT WAN



◆ Point of Presence

Network Protocol Overhead Impact

- **Simultaneous Optimization of CPU and Bandwidth Utilization**
- **Effective bandwidth is given by**
 - $B / (1 + O * B)$
 - $B = \text{bandwidth (bits/second)}$
 - $O = \text{protocol overhead (seconds/bit transmitted)}$



Protocol Support Limitations

- **TCP/IP Protocol Can Require Execution Power of Entire CPU of Y-MP8/864 for 700 Mbits/sec Bandwidth**
- **Effective Bandwidth Reduced 45%**



SAN DIEGO SUPERCOMPUTER CENTER

Wide Area Network Latency Can Require Data Buffering in Addition to Data Caching

- **Finite Speed of Light Creates Latency Between SDSC and LANL Comparable to Disk Seek Latencies**
- **Amortize Latency by Shipping Large Files**
 - $\text{Size} = 2 L * B$
 - $L = \text{Round-trip latency (seconds)}$
 - **For 800 Mbits/sec network, ship 8 MB files**



SAN DIEGO SUPERCOMPUTER CENTER

Integration of Local and Archival File Systems

- **Local CRAY Disk Supports 1000 Times as Much Data Transfers as Archival Storage at SDSC**
- **To Minimize Protocol Overhead**
 - Distinguish between
 - "Reliable" local file transport
 - "Unreliable" local network transport
 - Otherwise expect overhead to increase 20%



SAN DIEGO SUPERCOMPUTER CENTER

Integration of Local and Archival File Systems

- **User-initiated File Archival Storage Results In**
 - 1/7 of the data being archived
- **Automatic Migration of Local Files**
 - Allow non-uniform file migration across different file systems
 - /root versus /usr/tmp



SAN DIEGO SUPERCOMPUTER CENTER

Supercomputer I/O Scaling

- **For a 100 Gflops/sec Supercomputer**
 - Scale I/O by ratio of CPU speeds
 - Expect 3000 times as much I/O
- **Massive data generation may require dynamic regeneration of data rather than storage**



SAN DIEGO SUPERCOMPUTER CENTER

CPU Execution Scaling

Cache Level	Capacity (TB)	Data Rate (MB/s)
SSD	3	50,000
CRAY disk	126	50,000
Archive disk	210	450
Tape robot	6000	60
Shelf tape	12000	60



SAN DIEGO SUPERCOMPUTER CENTER

File Server Paradigm Shifts

- **Data Storage Requirements Will Be Increased by**
 - Integration of Local and Archival Systems
 - Higher Speed Supercomputers
- **Possible Shifts**
 - Local regeneration of data
 - Remote File Access



SAN DIEGO SUPERCOMPUTER CENTER

N93-13138

**MASS STORAGE SYSTEM
EXPERIENCES AND FUTURE NEEDS
AT
THE NATIONAL CENTER FOR ATMOSPHERIC
RESEARCH**

**Summary of the Presentation to the Conference on
Mass Storage Systems and Technologies for
Space and Earth Science Applications
July 23-25, 1991**

b y

**Bernard T. O'Lear
Manager, Systems Programming
Scientific Computing Division
National Center for Atmospheric Research**

3/5/91

This is a summary of the presentation given at the Conference on Mass Storage Systems and Technologies for Space and Earth Science Applications. The presentation was compiled at the National Center for Atmospheric Research (NCAR), Boulder, Colorado. NCAR is operated by the University Corporation for Atmospheric Research and is sponsored by the National Science Foundation. Any opinions, findings, conclusions, or recommendations expressed in this paper are those of the author and do not necessarily reflect the views of the National Science Foundation.

This presentation is designed to relate some of the experiences of the Scientific Computing Division at NCAR dealing with the "data problem." A brief history and a development of some basic Mass Storage System (MSS) principles are given. An attempt is made to show how these principles apply to the integration of various components into NCAR's MSS. There is discussion of future MSS needs for future computing environments.

NCAR provides supercomputing and data processing for atmospheric, oceanic and related sciences. This service is provided for university scientists and for scientists located at NCAR. There is a total of about 1200 users.

The data problem for this community can briefly be summarized as follows; Historical atmospheric data is archived, programs are saved and the data which model the atmosphere, oceans and sun are saved. The NCAR storage experience is based upon current supercomputing megaflop rates which produce a number of terabytes archived on a yearly basis. There is a history of data growth and file growth. The NCAR data storage experience has been as follows; There are about 500 bytes of information archived for each megaflop of computing. When NCAR had an X-MP/48, the archive rate for the utilized megaflop compute rate was 3 terabytes per year. The installation of a Y-MP8/864 increased the archival rate to 6 terabytes per year. Forecasting future computing configurations and atmospheric models being planned we are now approximating a 30-50 terabyte archive per year rate by the year 1993 or 1994.

Data has been saved in many forms over NCAR's existence and then migrated to machine-readable media. Some of the data has come from handwritten logs, from punch cards, half-inch tape. All of this has been collected and is now archived on IBM 3480 cartridge tape. One of the basic principles for archiving this data is to identify certain classes of data. Archive data is kept forever. Long-term

data is kept for 10 to 15 years. Near-term data is kept for 1 month to 1 year and a category called scratch data is killed after 1 month and cannot be recovered automatically by the system.

One of the other basic principles that has been identified is that dataset sizes continue to grow as a function of supercomputing sizing. The amount of data that can be saved is bound in storage by media capacities. That is, these criteria are established for determining which data will be saved and for how long because there is not an infinite media capacity at this time. Our experience has shown that every 10 to 15 years the data in the MSS will need to be migrated to a new media base because of changing systems and obsolescence of existing media. Usually the media or the drives cannot be purchased anymore. This migration takes place not because the data is bad on the media, but because the drives will not be available.

Another problem is that a number of companies have provided the capability for this massive storage, but the small companies tend to disappear within five years. The drive components that have been furnished for mass data storage disappear in five to eight years no matter what company they come from.

The next basic principle is that the migration of the mass storage system data to a new media base, which is now several ten's of terabytes, is not a trivial operation. The migration does not take place in a short amount of time. For instance, one-time migrations can run for long periods of time, necessarily years to move terabytes data. It is very difficult to guarantee that the data is migrated absolutely without reading it back, which is time consuming. These migrations are very costly and in my opinion shouldn't be done. We have developed the concept of "DATA OOZE," and we prefer this technique over migration right now. The way DATA OOZE works is that it is a continuous movement of data within the system. The data is moving across the

storage hierarchy and across the changing media types under the control of the MSS. The migration path for this data in the hierarchy can be from memory to solid state disk to high speed disk to disk arrays or farms, and from there out to some kind of tape. Later on as new data storage media become available, the data is migrated onto these media in real time, since every day some amount of the data is migrated as it is being used.

Our conclusions from these experiences have been that new components and media types are integrated according to the following rules; Use standard components. The standards may be real or de facto and apply in the areas of channels, interfaces, operating systems, media, etc. We look for media that is easy to obtain and is cost effective. We look for the long-term viability of the vendor and multiple sources for the many system components. In the area of mass storage system integration we look at access speeds, ease of expandability, heterogeneous host access, maintenance costs, media costs and systems costs.

There are a number of future growth issues for the NCAR MSS. The Scientific Computing Division (SCD) continues to develop future configuration scenarios. These scenarios try to anticipate the functional requirements we anticipate providing for our scientific community. There are three key components we need to address: network services and access, the large scale computing (Big Iron), and the data archives. Of course, these all play within the context of distributed computing.

The near-term issues for the NCAR MSS focus on some immediate upgrades which will deal with the MSS growth for a couple of years. The entire archive will be migrated onto double density 3490 and 3490-compatible media. The mid-90s to late 90s became more interesting because of the expanding interest in archiving vast data collections.

The issues of future growth will be centered in three areas of ongoing development: the various MSS software packages, the data storage components and the networks.

The questions then become how all of these components get assembled and which ones do we plan to use. Will SCD be able to construct an effective peta-byte MSS by the end of the decade? Which of our basic principles can we apply to insure that such a system can be built?

###

**Mass Storage System
Experiences and Future Needs
at
The National Center for Atmospheric Research**

**Conference on Mass Storage Systems
and Technologies for
Space and Earth Science Applications**

July 23-25, 1991

**Bernard T. O'Lear
Manager, Systems Programming
Scientific Computing Division
National Center for Atmospheric Research**

scd NCAR Scientific Computing Division
Supercomputing • Communications • Data

**The following presentation was compiled at the
National Center for Atmospheric Research
Boulder, Colorado**

- **The National Center for Atmospheric Research is operated by the University Corporation for Atmospheric Research and is sponsored by the National Science Foundation.**
- **Any opinions, findings, conclusions, or recommendations expressed in this talk are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.**

scd NCAR Scientific Computing Division
Supercomputing • Communications • Data

Introduction

The experiences of a scientific center dealing with
"The Data Problem"

- Brief history
- The current computing environment
- Development of some basic principles
- How the principles apply
- Future needs for future computing environments



History

NCAR provides supercomputing and data processing for atmospheric, oceanic, and related sciences:

- At universities
- At NCAR
 - Totals about 1200 users

The data problem for this community

- Save and archive historical atmospheric data
- Save programs and data which model the atmosphere, oceans, and sun

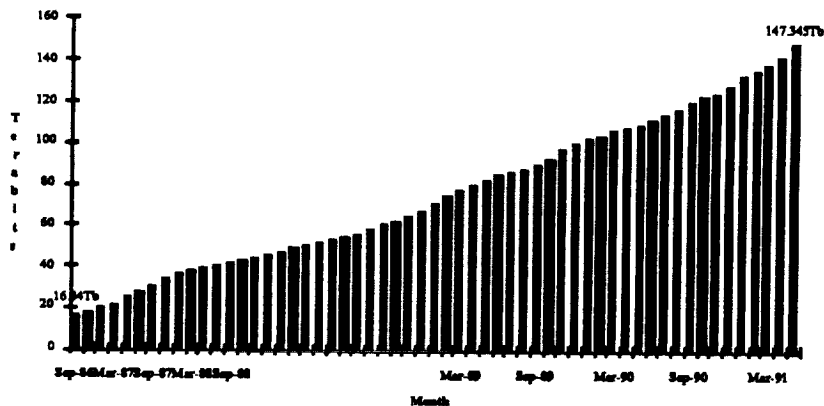


The NCAR Storage Experience

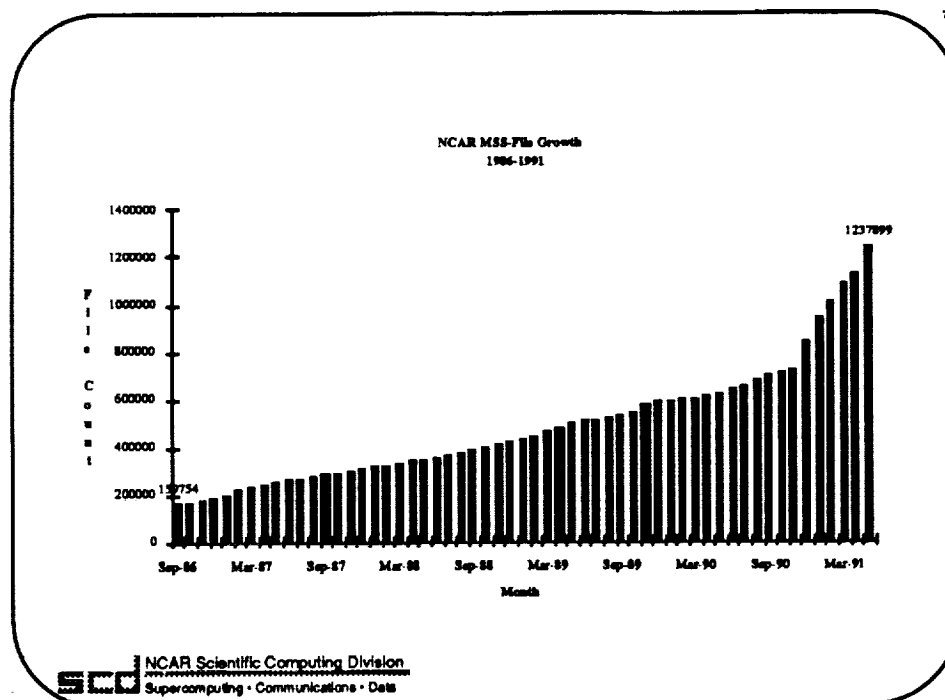
- 500 Bytes per million flop
- Archival rate for model output
 - 4 TBytes/year with X-MP/48
 - 8 TBytes/year with Y-MP8/864
 - 40 TBytes for climate simulation

NCAR Scientific Computing Division
Supercomputing • Communications • Data

NCAR MIP-II Data Growth
1986-1991



NCAR Scientific Computing Division
Supercomputing • Communications • Data



**NCAR Mass Storage Systems
(MSS)**

Usage Data

101,700 tape cartridges in use
Over 18.5 Tbytes of data stored
Over 710,000 files
Average file length 26.2 MB

Fast Path


≤ 2 minute delivery

NCAR Scientific Computing Division
Supercomputing • Communications • Data

History (Continued)

Data saved in many forms -- then migrated to machine readable media:

- Handwritten logs -----> Punched cards
- Punched cards -----> One-half inch tape
- One-half inch tape -----> AMPEX TBM tape
- AMPEX TBM tape -----> IBM 3480 tape
- IBM 3480 tape -----> IBM 3490-E tape
- IBM 3490-E tape -----> ? ? ?

 NCAR Scientific Computing Division
Supercomputing • Communications • Data

Basic Principles

Identification of data classes:

- Archive data = keep forever
- Long-term data = keep 10-15 years
- Near-term data = keep 1 month to 1 year
- Scratch data = kill after 1 month

 NCAR Scientific Computing Division
Supercomputing • Communications • Data

Basic Principles (Continued)

- Dataset sizes continue to grow as a function of supercomputer sizing
- Dataset sizes are constrained in Storage by media capacities
- Every ten to fifteen years, the data in the MSS will need to be migrated to a new media base

Basic Principles (Continued)

Migration:

- Not because the data is bad on the media...
- But because the drives will not be there
 - Half life of a start-up company = 5 years
 - Half life of drive electronics = 5-8 years

Basic Principles (Continued)

The migration of MSS contents ("n" tera-bytes) to new media
is not a trivial operation

One-time migrations:

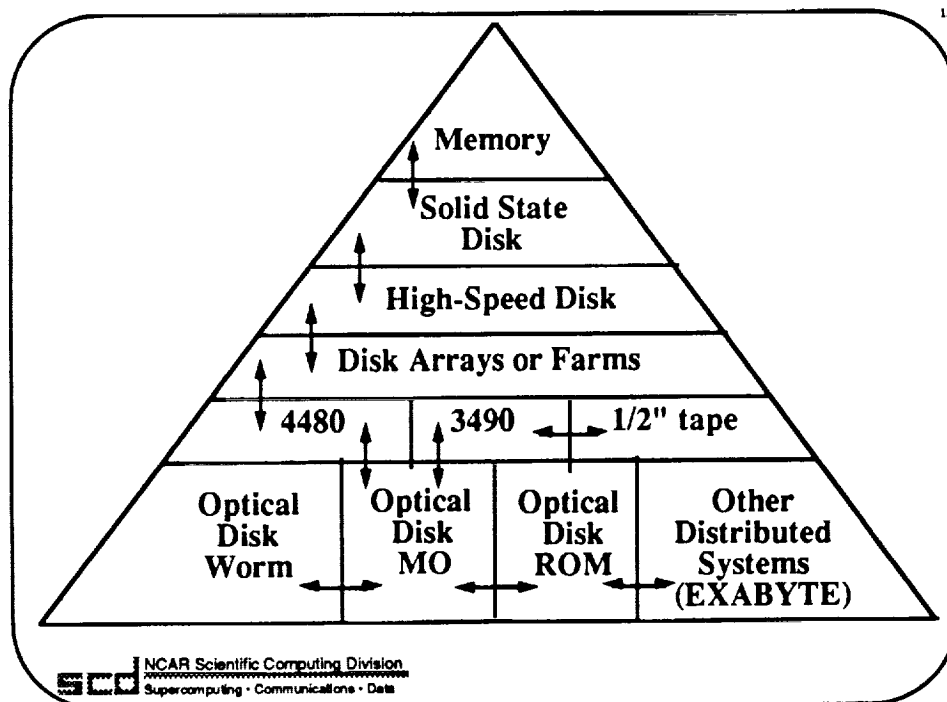
- Run for long periods of time (years)
- Are difficult to guarantee
- Are costly
- Shouldn't be done

Basic Principles (Continued)

Data OOZE preferred over migration

Data OOZE is a continuous movement of data within the system:

- Data movement across:
 - The storage hierarchy
 - The changing media types



16

CONCLUSIONS

New components and media types are integrated according to these rules:

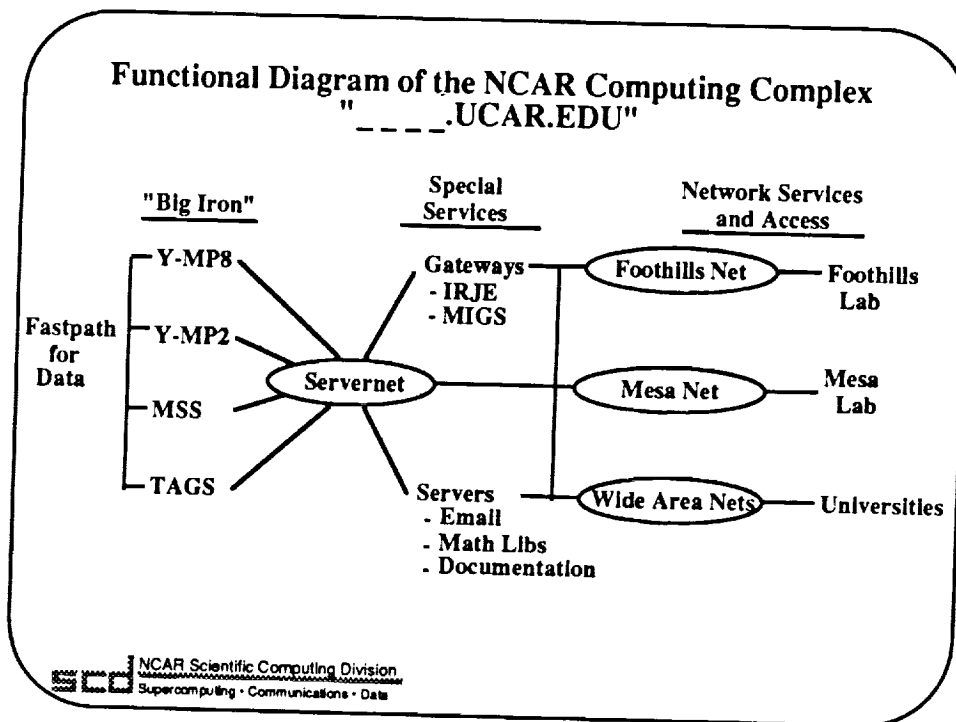
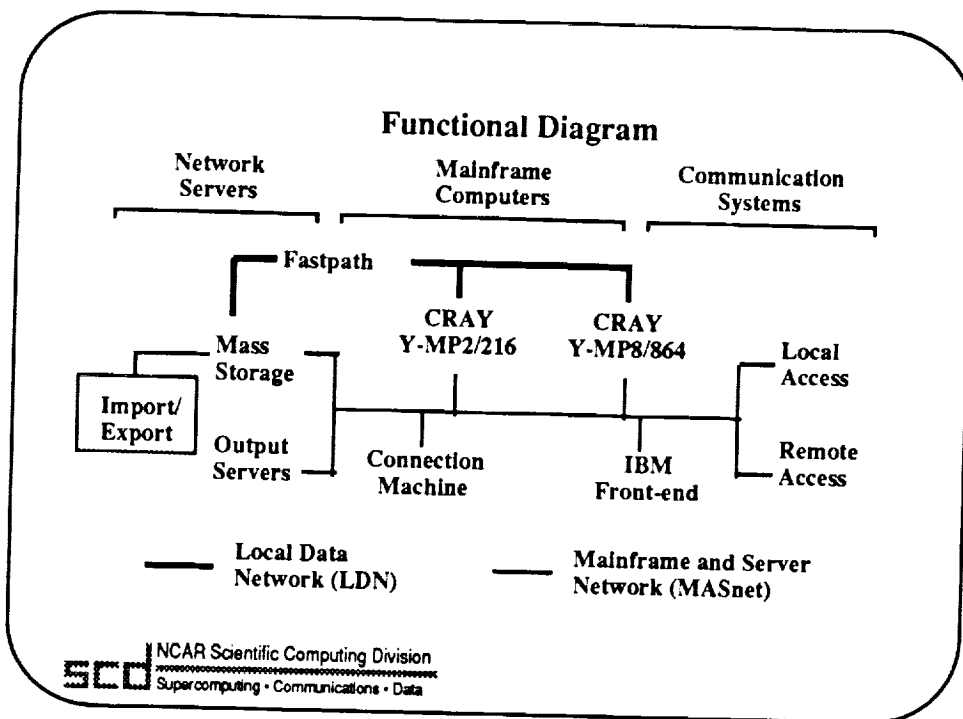
- Standards (real or de facto)
 - For channels and interfaces (IBM, IPI, HIPPI, SCSI)
 - For media
- Long-term viability of vendor
- Multiple source availability for media (drives?)

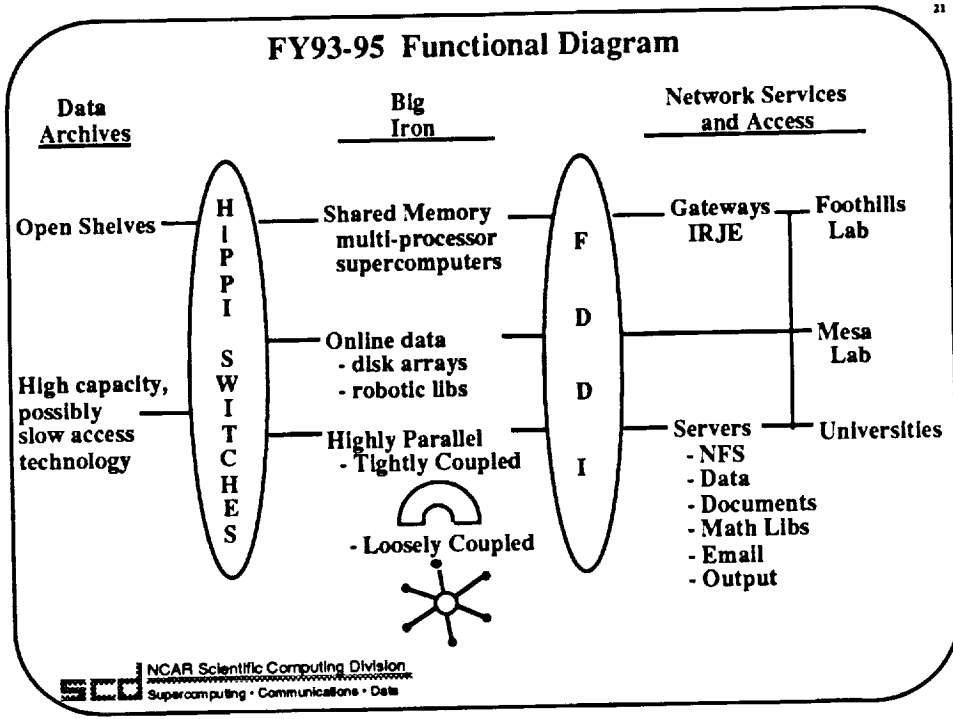
SCD NCAR Scientific Computing Division
Supercomputing • Communications • Data

MSS Integration

- Access speeds (sometimes)
- Ease of expandability
- Multiple heterogeneous host access
- Maintenance costs
- Media costs
- System cost

FUTURE GROWTH ISSUES IN THE NCAR MASS STORAGE SYSTEM





NCAR MSS Near Term Upgrades

1. Purchase (IBM) 3490E drives for double density capability
2. Automatic double density migration takes place for shelf archive
3. Hope is STK furnishes double density for drives on ACS in < 6 months.

NCAR Scientific Computing Division
Supercomputing • Communications • Data

NCAR MSS

The Issues of Future Growth are dependent upon:

1. Future MSS Software
 - a. Distributed MSS
 - b. Large archives (Peta-Byte)
2. Future Data Storage
 - a. The media
 - b. The drives
 - c. The robotics
3. The Network and Channels
 - a. HIPPI
 - b. Fibre channel standards
 - c. fabric (switch)

1. Future MSS Software

- a. Distributed MSS
 - UNITREE (DISCOS)
 - Infinite Storage Architecture (EPOCH)
 - Distributed Physical Volume Repository (EPOCH & STK)
 - EMASS (E-SYSTEMS)
 - NAStore (NASA, Ames)
 - NETARC and AWBUS (CDC)
 - SWIFT (IBM)
 - DataMesh (Hewlett Packard)
 - M (DS) ² NASA Goddard
- b. Peta-Byte Archives
 - How do we build them?

2. Future Data Storage

a. THE MEDIA

- The 3M National Media Laboratory (Media Database)
- Government funds and goals
- Private sector participation
- Standards being developed

b. THE DRIVES

- Being developed for the media
- 10-year life span
- Attachable to various robotics

c. THE ROBOTICS

- StorageTek is the leader
- ODETICS
- EXABYTE and others

3. The Network and Channels

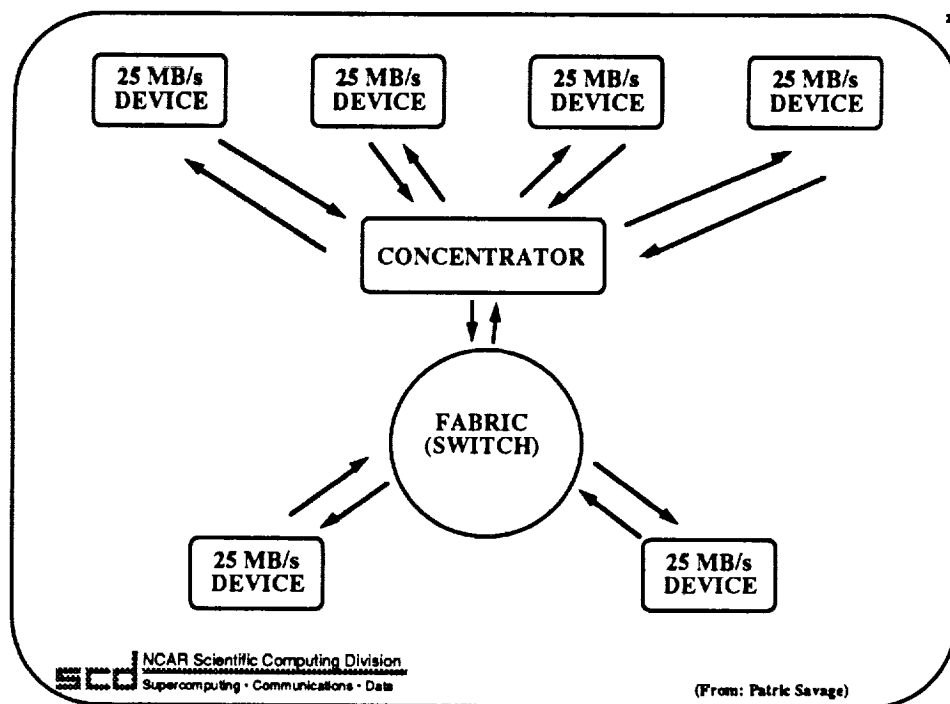
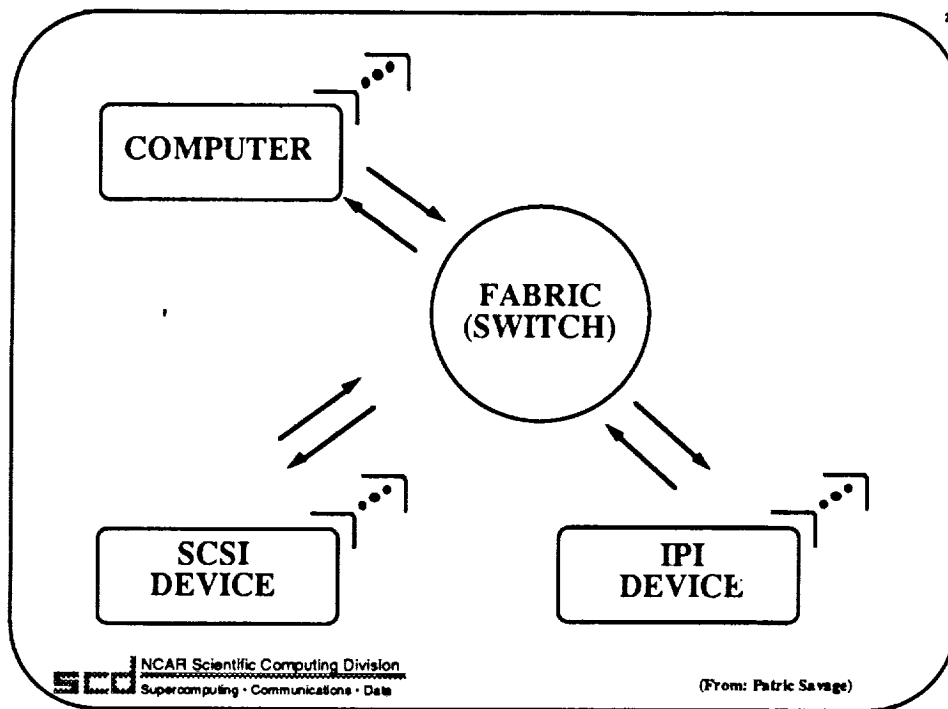
a. Standards moving fast for HIPPI

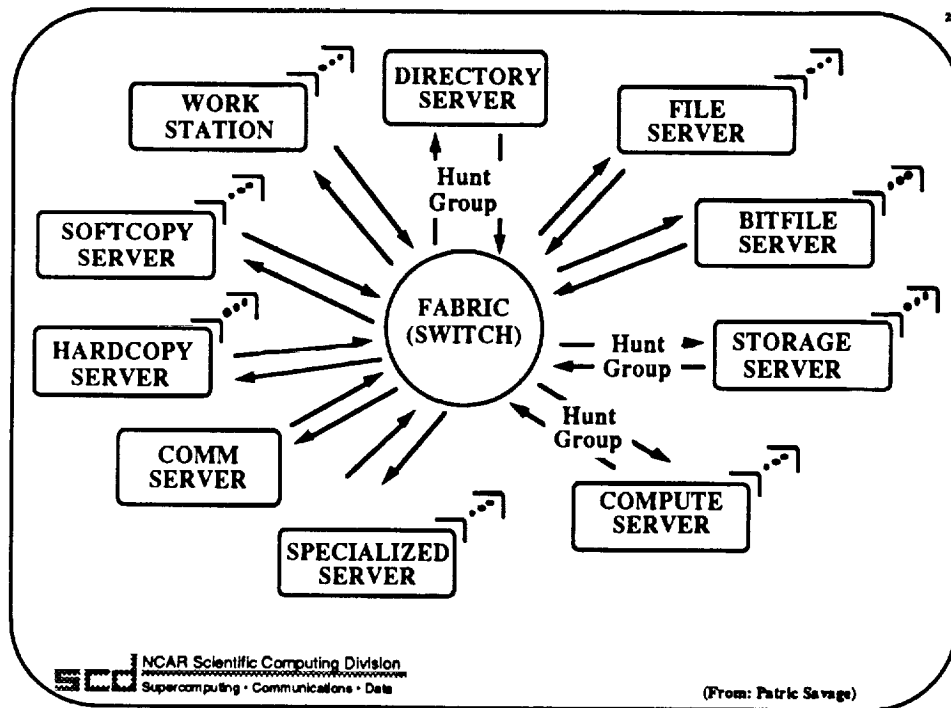
- The HIPPI switch

b. Fibre Channel Advantages

- Length to 10 kilometers
- General Protocol
 - HIPPI
 - SCSI
 - IPI
 - Others
- Security
- Immune to Electrical Disturbance

c. Fabric Switch





The Long Hold: Storing Data at the National Archives

Kenneth Thibodeau, Ph. D.
 Director, Center for Electronic Records
 National Archives and Records Administration

The National Archives is, in many respects, in a unique position. For example, I find people from other organizations describing an archival medium as one which will last for three to five years. At the National Archives, we deal with the centuries, not years. From our perspective, there is no archival medium for data storage, and we do not expect there will ever be one. Predicting the long-term future of information technology, beyond a mere five or ten years, approaches the occult arts. But one prediction is probably safe. It is that the technology will continue to change, at least until analysts start talking about the post-information age. If we did have a medium which lasted a hundred years or longer, we probably would not have a device capable of reading it.

The issue of obsolescence, as opposed to media stability, is more complex and more costly. It is especially complex at the National Archives because of two other aspects of our peculiar position. The first aspect is that we deal with incoherent data. The second is that we are charged with satisfying unknown and unknowable requirements.

The data is incoherent because it comes from a wide range of independent sources; it covers unrelated subjects; and it is organized and encoded in ways that not only do we not control but often we do not know until we receive the data.

The sources are potentially any operation of the Federal Government, or its contractors. The National Archives has been in the business of collecting digital data for two decades. The way we get it is through our authority over all federal records. Under the Federal Records Act, no agency of the Federal Government can destroy or alienate any Federal record without authorization from the Archivist of the United States, who is the head of the National Archives and Records Administration. Simplistically, the way it works is that agencies tell us what records they have, and we tell them which ones they can destroy when they no longer need them, and which ones must be preserved for posterity. (The definition of Federal record in the law explicitly includes machine-readable files.)

Since 1972, we have reached agreements with agencies that provide for them to transfer to us, and for us to preserve, data from 600 data collections. 573 of these are still active. From these agreements, we have received over 10,000 data files. The rate of transfer has increased dramatically in the last two years: In fiscal year 1988, the National Archives received 167 data files. So we are currently operating at eight times the volume of new files we had years ago, and we expect at least to double next year.

Those numbers are very encouraging, but the overall picture is rather bleak. If we look at all of the data which was scheduled to arrive in the last twenty years, from those 600 data collections, we have received less than 7% of the transfers which should have been made. We recently completed development of a system to generate dunning letters to agencies who fail to transfer data as scheduled, and to track each case to completion. But this system creates additional problems. If I implement it as planned, on a government wide basis, we would need to increase our capability to handle new files, not by doubling current capacity, but by increasing it more than six times. And to handle the backlog of data which should have come in before now, I would need at least 10 times our current capacity.

The past gives us pause. But the future is a brave new world. At least it requires a degree of bravura just to glance in that direction. We have underway a study which is looking beyond the 600 data collections we have decided to preserve to see what else is out there. It is a study of major federal databases being conducted by the National Academy of Public Administration

PRECEDING PAGE BLANK NOT FILMED

(NAPA). This study has some interesting exclusions. First of all, we told NAPA not to bother with systems used for generic housekeeping functions, such as personnel, payroll, procurement and supply, because there is little likelihood that we would have any interest in preserving data from such a system. Secondly, we told them not to look at big science, because that is such a large and complex area that it deserves separate attention. (We hope to engage in a project with the National Academy of Sciences on the preservation of scientific data.) Thirdly, we told NAPA not to worry too much about databases on PCs, simply because they would never finish the project if they tried to find all the interesting databases sitting on desktops. With those limitations, NAPA has identified over 10,000 databases.

Obviously, that is far too big a number even for us to think about. So we gave NAPA a set of criteria for culling from the total inventory a subset of those databases with some likelihood that the National Archives would be interested in preserving them. We thought we might wind up with a list of the 500 most important databases in the Federal Government, from an archival perspective. That list would pose quite a challenge for us, because it could practically double the total number of data collections generating data that we want to preserve. The subset of 500 currently has about 900 members.

The next phase of this study is to solicit advice from subject area experts about what data we should try to preserve. NAPA has organized five working groups, with a total of 32 experts in a variety of fields. We are bringing these people together at the end of July for a four day meeting where they will try to develop some common opinions on the long term value of the data.

Which brings me back to the basic point here: what we are dealing with is incoherent data. It concerns practically any area in which the United States Government is involved, which is practically anything. The data we already have ranges from data about tektites on the ocean floor to military operations in time of war. It includes census data on population and the economy, data on Japanese-American internees in World War II, detailed data on air traffic and on stock and bond transactions, and on many, many other subjects. The variety of subjects covered is also increasing.

The data is extremely diverse in content, but content is often the only thing we know about the data until it comes in. We know how many transfers are due, but most often we do not know what the volume of data in a transfer will be, or how it will be organized, even at the physical file level. For example, the files which came in during the first six months of this fiscal year ranged in size from 6 K to 1.4 gigabytes. The number of files in a transfer has ranged from one to 400, and we expect some transfers in the next few years will contain thousands of files.

One thing we do know about the data before it arrives is its logical structure: everything we receive is in flat file format, because we require it to come in that form. However, we realize that this requirement is unreasonable and unrealistic in many cases. We are working to expand the range of formats we will accept to include relational tables. We expect to change our regulation to that effect by the end of this year. We know that, when we do that, it will be only one of many steps we will have to take in a journey with no foreseeable end.

That is a brief overview of one aspect of the unique situation of the National Archives. The second aspect is that we are charged with satisfying unknown and unknowable requirements.

NARA's mission to preserve and provide access to records with enduring value makes NARA, in effect, the agent of generations yet unborn. What differentiates this agency from other parts of the government is the unique responsibility NARA has to serve the information needs of the distant future. This responsibility is fundamental to the very essence of the National Archives as keeper of the Nation's memory.

NARA's responsibility to the future places us in a perpetual quandary: we must devote ourselves to serving needs which we cannot know. We cannot know the questions the future will ask of its past, nor how future researchers will go about answering these questions. We must assume, however, that the information technology which will be available in the future --

- even in the very near future --- will be more powerful and more flexible than what is available today. Information processing problems which today are difficult and costly, if not impossible to solve will become as simple as getting a computer to print out narrative in paragraph form. (A short 20 years ago that was beyond state of the art.)

Along with the technology, analytic tools will continue to improve: there will be further developments as powerful as the mathematics of chaos which will help researchers to understand things which today appear to defy reason. We can also assume that events will happen in the future, which will be as threatening as the depletion of atmospheric ozone, or as exciting as Operation Desert Storm, or as commonplace as the passing of generations, which will make future users want to go back to reexamine the records of the past.

EVENING SESSION

(8:30 p.m.)

DR. HARIHARAN: Ladies and gentlemen, it is my pleasure again to introduce Dr. Mallinson, who will be giving us a talk about his reminiscences in the field of magnetic recording over the last 40 years.

(Laughter)

DR. HARIHARAN: Dr. Mallinson has an M.A. Degree in Natural Philosophy in Physics from University College in Oxford, England. And he joined the Ampex Company in Harrisburg, Pennsylvania in 1954 to work on the theory and design of magnetic lodging elements.

In 1962, he joined Ampex Corporation in Redwood City, California, where he held many positions concerned with the understanding and development of magnetic recording systems. From 1976 to 1978, as Manager of Hybrid Magnetic Recording in the Data Systems Division, he was concerned with the initial design of the 750 MBS digital recorder.

From 1978 to 1984, he supervised the Magnetic Recording Technology Department Multidisciplinary Group, working in magnetic recording theory, high density head fabrication, coding and communication theory, and the exploration of advanced concepts in various areas of recording. In 1984, he was appointed as the Founding Director of the Center for Magnetic Recording Research at the University of California, San Diego. Since 1990, he has been the President of Mallinson Magnetics, Inc. He has published over 60 papers on a wide variety of topics in magnetic recording.

Dr. Mallinson was an IEEE Magnetic Society Distinguished Lecturer in 1983. In 1984, he was awarded the Alexander M. Poniatoff award, named after the founder of Ampex Corporation, an award for leadership in the theory and practice of magnetic recording. Dr. Mallinson?

(Applause)

BANQUET PRESENTATION

Professor John C. Mallinson

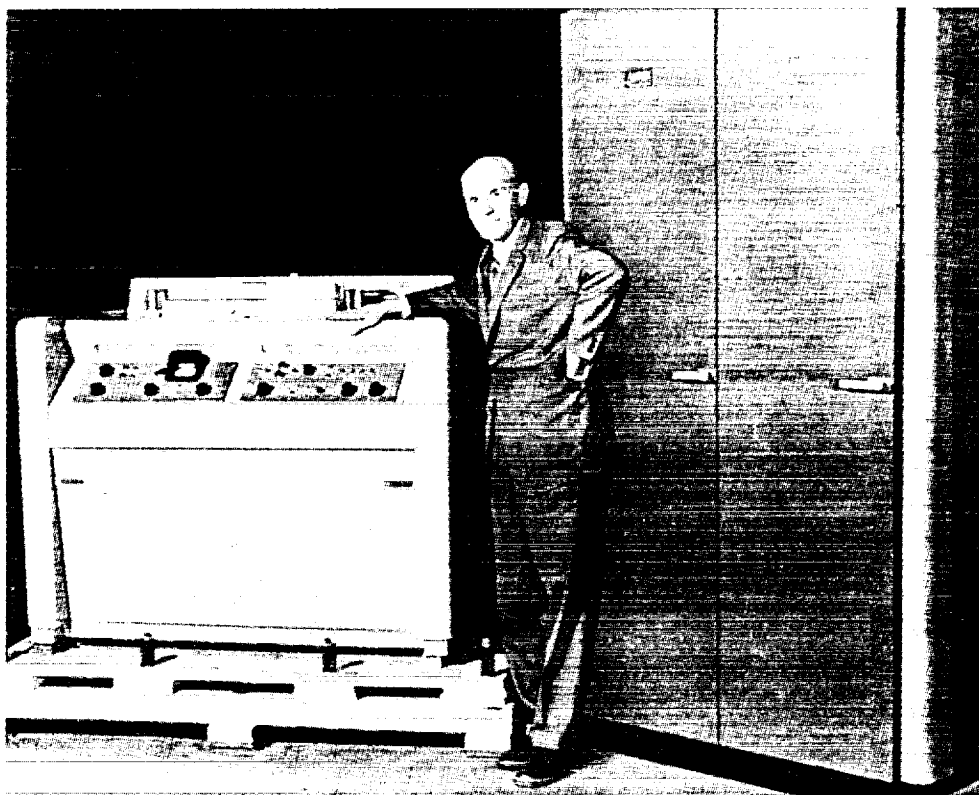
DR. MALLINSON: Thank you, Hari I've given a great deal of thought to what I should say to you this evening. Options ranged from telling you a number of risque jokes; but I realize that I have, in fact, been invited to give some reminiscences on--it's not a 40-year career; I'm not that old --

(Laughter)

DR. MALLINSON: But I have had a career since 1962 in magnetic recording. The most important of those years, I think, were the years at the Ampex Corporation; I was there for 24 years.

So, basically, what I'm going to talk about is some reflections on 24 years at Ampex.

First of all, I want to tell you that I think in the field of magnetic recording research and development, Ampex Corporation has very, very few equals. I won't list for you all the things that Ampex has developed; but most prominent amongst them is surely the invention of video tape recording.



ORIGINAL PAGE
BLACK AND WHITE PHOTOGRAPH

DR. MALLINSON: There is a picture of Alex M. Poniatoff, the founder of Ampex, standing beside a video recorder. Alex M. Poniatoff hired me in 1962 to join the Ampex Corporation, and we hit it off together almost immediately because I was a pilot in the British Royal Air Force and still am a pilot, Alex M. Poniatoff was a pilot in the Russian Air Force in 1919.

And I will recount for you one story of Alex M. Poniatoff, and you should just think about this story while looking at Poniatoff when he is 76 years of age there.

In 1919, Alex M. Poniatoff was a Captain in the Russian Air Force. He was flying a wooden six-engined aircraft; I'm not sure what it was. Let's call it a Sigorsky; and they were flying it off some lake somewhere in Russia.

And one day, he went out to the lake and noticed that the red flag was flying; and he knew, as a good Captain in the Russian Air Force, that that meant that the commanding officer had decided that today the water was too rough for flying.

But Alex M. Poniatoff thought otherwise, as a young "Turk," as they say, in the Russian Air Force. He gathered up his copilot and his engineer, and they commanded someone to row them out to the flying boat; and they fired up the engines, turned it into the wind, and commenced the take-off run.

But I am not telling this story anywhere near as funny as he told the story; but long before the thing even got on the stack, before it even started showing any signs of becoming airborne, this wooden flying boat disintegrated. It broke up.

And shortly thereafter, he found himself in this icy cold water, calling out for his copilot and his engineer; and they all swam to each other and must have thought: What the hell are we going to do now?

And they noticed another small boat; this time it was a motor boat coming out, with their commanding officer in it. And he told me this hilarious story about what the commanding officer had to say about him; it was the end of his career flying the flying boats in the Russian Air Force.

So, that is Alex M. Poniatoff standing in front of the first Ampex video recorder in 1956. The video recorder is that unit; that unit weighed 1,100 pounds. And behind him is this cabinet which had 275 vacuum tubes in it; that was the electronics of the thing.

A reel of tape on that machine was 12 inches in diameter, 2 inch tape, and it recorded black and white television--NTSC television--for one hour. That was in 1956.

Before 1970, people had decided to use this video recorder for recording digital information; and such a reel of tape held 30 times 10^9 bits of information. I'll try and stick with bits all the time. Video people talk bits; computer people call big bits bytes.

They had 30 times 10^9 , 30 gigabytes, of data on it. And Ampex at that time, as was mentioned in the session today, started selling a system called the Terabit Memory, the 10^{12} bit memory. A 10^{12} bit memory had no less than 32 of those units.

And basically, what I'm going to talk about-- after some remarks--is how the evolution of recording since, say, 1960 to 1990, over the last 30 years, has resulted in much more compact ways of recording 10^{12} bits of information.

At that time at Ampex, when these terabit systems were being made, with their 32 transports and their four Philco computers controlling them, we used to laugh at the idea of a terabit in a drawer; it was a joke. Well, a terabit in a drawer is now something which is available to everyone. I will show you a foil of it later on.

Rather than show foils all the time, which would just make it into a technical talk, I thought I would just plain talk to you about one of my major reflections at Ampex; and that is how it is that every two or three, perhaps every five, years it was perceived, though, that there was some important threat to the dominance of magnetic recording. And I'm afraid to tell you that that still goes on today, this notion that there is something around that is going to displace magnetic recording.

After 30 or 40 years of it, I very, very much doubt it, but I am getting ahead of myself.

The first thing that came along in the early 1960s, probably because so many of the Ampex engineers were involved in television and television revolves around cathode ray tubes--TV tubes--the notion came up that an electron beam recorder would be the best thing. It would be the answer to the maiden's prayer.

And Ampex, in the early 1960s, made electron beam recorders. The electron beam recorders recorded on photographic film that was specially made for us by Kodak; the track following servo was done by wobbling the electron beam, the dithering method of track following and observing scintillations of the scintillator coating on the back of the photographic film.

It recorded 100 megahertz analog bandwidth, which supported two 30 megabit per second channels. It was considered to be something wonderful in 1960 because the terabit memory was only about 5 megabits per second per channel, and this electron beam recorder had two 30 megabit per second channels.

What was wrong with the electron beam recorder like that? First of all, everything had to be done in a vacuum. Electron beams don't go too far in the air.

Secondly, it used photographic film, which had to be taken out of the vacuum and developed and put back in the vacuum to play back.

And thirdly, it was an enormously large machine. The electron column on it looked like a regular electron microscope. It was about 8 feet tall; and in fact, the customer for the electron beam recorder, the U.S. Air Force, found it necessary to increase the size of the cargo hatch of the C-130 at the time. It just wouldn't go in.

So, that was one of the things that had us all abuzz. Electron beam recording was going to solve all the problems. In retrospect, it's hard to know why people were so naive to think such a thing as that.

Shortly thereafter, a company was started in Boston; one of the principals, Dennis Spiliotis, is here. It was called Macrobit. Macrobit had a similar notion that they were going to use electron beams to record on silicon wafers. Single crystal silicon had more or less just become available in the mid-1960s, and their notion was basically to lay out photolithographically little capacitors on a silicon wafer--tell me if I'm wrong, Dennis.

They were going to lay out little capacitors on a one-inch square, or maybe a one-and-a-half inch square, single silicon wafer; and it would have a capacity of about 10 megabits. And that one failed, I believe, because 10 megabits is a ridiculously small capacity.

At least at Ampex, it was realized that since Ampex's principal business was recording massive amounts of analog signals or data, 10 megabits was just too small a module to be of interest.

So, electron beam recording on silicon disappeared.

There was also a slight cautionary tale about that microbit memory in that its capacity did not exactly match the then-IBM disk drives. The IBM disk drives of the period were 3330s; and the Macrobit memory had something like 7 percent more capacity than a single 3330, which in some spheres was its death knell because it's one thing to say that we will have an excess capacity in some memory device; and if we don't need to use all of that, we will just fill it in with garbage--you know, we'll just add noise, random 1's and 0's.

It is one thing to say that; it is altogether another thing to find some software engineers or some people who will do that. So, there's another lesson to be learned.

The two main reasons for electron beam recording were, of course, that you could do the recording without touching the medium. Touching the medium or not touching the medium was considered to be an extremely important fact; and that, in turn, led to the idea that there should be optical recorders of one kind or another.

I hate to think how much money the Ampex Corporation dumped into optical recording. Magneto-optical recording, which has been talked about this morning by the man from Alphatronix, is in its third--what's the right word?--life at the moment.

Originally, in the 1960s, the first life--the first period--of magneto-optical recording was being done on manganese bismuth; that was the material. You had to put in an awful lot of power on the manganese bismuth to heat it up to its Curie temperature of 400⁰ or 350⁰ centigrade.

And then, the material basically failed because, after repeated cycles to that high temperature, it changed phase; and the Curie temperature changed. It was unstable material; it would not stand indefinite recycling from room temperatures up to 400⁰ centigrade.

Now, the second phase of magneto-optic recording led by Big Blue, was of course the Europium oxide run, which had terrible trouble that its compensation temperature was at liquid nitrogen temperatures. It failed. And the current one, which is iron cobalt--some rare earth--usually terbium--may or may not make it. Mr. Freese this morning said it would make it in a "niche" market; and I think that's about the right way to think about it.

What was the perception at Ampex of magneto-optic recording? Why, it was that Ampex should go ahead and make a magneto-optic tape recorder. And in 1966, there were shipped to the U.S. Navy no less than five magneto-optic tape recorders.

These tape recorders had 2 inch wide tape in them; the recording was transverse scan, just like the quad video machines. The transverse scan was done with a rotating polygonal mirror that was going round at 14,400 rpm, just the same speed as the quad drums. They were writing on a hard magnetic film with a permalloy overlayer and the readout scheme supported a data rate of 10 megabits per second.

We should have been smarter at that time to realize that this was in the mid-1960s; 10 megabits per second was already much slower than magnetic recorders were going in the mid-1960s.

But nevertheless, the hope persists in the human animal that there will be some other technology that will get you out of your perceived present problems.

The magneto-optic recorder was expensive; and I forgot to mention that, at that time in optical recording, there were no gallium arsenide solid-state diode lasers around. The lasers were all helium-neon gas lasers; and the gas lasers in this magneto-optic recorders shipped to the Navy were fully 24 inches long--helium-neon gas lasers. So, that failed.

I would say that basically the principal reason it failed was that it had material problems; and in particular, it had material problems where it was not possible to get more than 10 megabits per second data rate through it.

And that, I believe, is a persistent problem that is related with magneto-optic recording throughout all of its history including today. It is hard to get high data rates. So, it is a question to do with signal-to-noise ratio, which we don't have to go into.

50 megabits per second today is considered a high data rate for magneto-optic recording of any kind.

At the same time, or shortly after we realized that the magneto-optic recorder was not quite the right way to go, there was work undertaken in WORM, laser melting or oblatting of some material like tantalum. It was very rapidly discovered that --

You know, the initial discovery phase of all of these endeavors is very exciting and goes very rapidly; and it is only later that we realize that you haven't got anything worthwhile.

In the case of the laser oblatting, the pure WORM optical disk, it was very rapidly discovered that the laws of diffraction--Lord Rayleigh's laws of diffraction--do not apply in the writing phase. The writing phase is highly nonlinear. There is nothing that Lord Rayleigh ever said that precludes you from recording, say, one-tenth micron spots or 500 Angstrom spots.

And if you look back in the literature, in a journal called **SPIE**, the Society of Photographic and Instrumentation Engineers, you'll find papers from good old Ampex about recording 500 angstrom diameter holes which corresponded to an area density of nearly 10^{10} bits to the square inch.

The hooker comes when you realize that what Rayleigh's diffraction limit--the area of the spot on the disk and all of that diffraction stuff--applies with a vengeance on playback. You can't ignore it on playback.

And the Airy disk is just like the gap loss function in magnetic recording. The only way to make the gap loss function in magnetic recording smaller or shorter is to make smaller gaps. The only way to make the area disk smaller in optical recording is either to think about numerical apertures that are ridiculous, more than unity, or go to shorter and shorter wavelengths.

At that time, using gas lasers, short wavelengths were available; and so, there were experiments done with blue lasers. And that possibility doesn't seem to be on at the moment with gallium arsenide. Gallium arsenide has an energy gap of about one electron volt, and that means it puts out photons that are around 8,000 Angstroms in wavelength--800 nanometers.

And there are not many semiconductor materials with larger energy gaps than that. A blue one at 4,000 Angstroms would need a 2 volt energy gap for instance.

So, we worked on WORMs. The basic trouble with a WORM is that it's a small module. A WORM disk of reasonable diameter, 6 inches or 8 inches in diameter, is only going to hold a gigabyte or so of data.

In the framework of Ampex thinking, that is nothing. That won't support digital video for very long at all. Digital video recording--the standards for it--were beginning to be set in the late 1970s; and Ampex, in fact, made the world's first digital video recorder.

It was a parallel access disk recorder, and it ran at 84 megabits per second. It was a composite video being sampled at four times the color subcarrier.

And so, it was realized that making a WORM disk, an optical disk, with perhaps 10^{10} bits of information, wasn't going to support video for very long. So, that was canned.

The ROMs, the CD-ROMs, and the audio compact disk were never considered at Ampex because they are best regarded as a publication means. They are just a replication; it's a way of pressing disks to circulate information.

More or less at the same time, another pervasive idea in optical recording came up, which is still going to this day. There is a well-known laboratory in Texas that is still promoting the idea of doing holographic optical recording in three dimensions on crystals.

Ampex, of course, worked in that; the crystal, as always, was strontium niobate or lead niobate.

And I could go on for some time, if I wanted to make it a technical talk, about what's wrong with that idea. One of the things that is wrong with this holographic thing is that it depends on extreme mechanical precision in doing the optics.

It is not for nothing that when you go into an optic lab that you see everything is being done on granite blocks that are on air-bearing legs, and it's in an air-conditioned room. Holography works just fine, sending multiple beams through objectives. Lord knows in a telescope, or in this thing, whatever number you want to say, I'll agree with it. A million beams go through that lens.

But keeping them all in focus, keeping them all in the relative positions correctly, requires a dimensional stability and a vibration-free environment that makes it unlikely that it will work in any condition other than an optical bench in an air-conditioned room.

I'll tell you a little joke about optical recording because after-dinner speakers are supposed to tell jokes. It was told to me by a Dutchman, and the joke goes:

Do you know how you make a small fortune in optical recording? The answer is: You are either Edward Rothschild, who every year teaches courses with names like technology opportunity conference or something, or else you start with a large fortune --

(Laughter)

DR. MALLINSON: And that is exactly the way it has been in optical recording. Ask Schlumberger, ask Honeywell, ask Storage Technology.

So, what have I got left? I've got up there to about the mid-1970s; magneto-optic recording is coming around again in this third reincarnation in the oxide phase. Bubbles started to appear in the mid-1970s. Bubbles were, in my opinion, a loss leader right from the very beginning. The claim made by Andrew Bobeck of bubbles was that you would be able to get one million bubbles to the square inch and shift them at one million shifts per second.

Unfortunately, to do that, you had to have an entirely new material, a garnet substrate; but all my materials scientist friends told me it was going to be considerably more expensive to make than a silicon single crystal. That's overlooking the fact that single crystal silicon substrates were already available.

And then, if you looked at the structures that were involved in bubble recording, they were considerably more complicated; more area was required per bit cell. And bubbles have never gone anywhere.

There are still Japanese companies working on bubbles; I believe Hitachi still works on it. We at Ampex had a large bubble program because it was imagined, incorrectly, that bubbles were somehow going to do something for recording.

It's this constant notion that there is something there--some new technology--that hasn't been thought through very carefully; but it's new--new, new, new. New is the name of the game in research. That it would do something that semiconductors had not managed to do. I answered a question this morning about semiconductor memory; and believe it or not, good old, long-suffering Ampex had a semiconductor division down in Santa Monica in the L.A. Basin, and it was dedicated to trying to make a semiconductor memory.

And for those of you who weren't there this morning, the basic trouble with semiconductor memories is that there seems to be no chance that they are going to ever be economically worthwhile.

Magnetic recording, over the last 30 or 40 years, has continued to double in density every two or two and a half years.

I put forward the view this morning that the main reason that it was growing at that rate for such a long period of time was that that was the rate at which the electronics industry or the computer peripherals industry could accept technological change.

It just so happens that, apart from glitches to do with the dumping of semiconductors from you know where, the semiconductor industry has been doubling in density. They like to say every two years. In other words, it is following just about the same slope.

A gigabyte of semiconductor memory at the moment costs you \$500,000. A gigabit of magnetic memory in hard disk costs you \$1,000. So, there is a 500 to 1 differential in price; and the two technologies are advancing down the same sort of maturity curve.

If you take literally the slight differences in slope, then you come to the conclusion that they will reach equality in price in the year 2007. Now, that assumes that semiconductor technology can continue to advance in density at its current rate and so can magnetic.

In semiconductor technology, there seem to be some enormous barriers to do with good old diffraction of light coming up. Once you get down to quarter micron lines, it seems unlikely that even ultraviolet extreme blue light diffraction will do it. And you will have to go to X-ray lithography.

So, I don't believe that for a long time the semiconductor is going to be a threat to magnetic recording. It took Ampex a long time to work that out.

What in fact has happened, looking back with 20/20 hindsight--a retrospective look at things--is that every time there has been some advance in LSI, LSI has become larger, cheaper, denser, whatever the criterion is that you like to use--what has happened is that it has enabled some other form of recording to move to a higher level of performance.

At the Ampex Corporation, not only were rotary head video recorders invented in 1956, but helical scan recorders were invented two years later, in 1958. The helical scan recorder was invented because it had a long enough swipe down the length of the tape to get one whole television frame in--field--I beg your pardon.

Unfortunately, the time base errors of the head tape interface were much too high to be handled by the time base correctors that were in the early quad machines. The early quad machines--the very first ones--had mercury delay lines. The later ones had quartz crystals, where the delay time was changed by altering the voltage--the piezoelectric delay lines.

And that could not be used for the large timing errors in a long helical scan. Helical scan machines languished in the lab as an idle curiosity for almost five years, until, all of a sudden, I forget whether it was 8K or 16K DRAMs came along, that suddenly made it possible to do the time base correction digitally.

So, it was silicon technology that enabled helical machines to be useful. Likewise, in an optical disk at the moment--take the audio compact disk--the raw bit error rate coming off the disk, meeting the Philips or the manufacturing specification, is 1 byte--I shouldn't say byte; I promised I wouldn't say byte--let's say 1 bit error in 1,000, 1 in 10^3 , 1 in 10^4 . In order to get the error rate satisfactory for audio, which is 1 in 10^{10} , which is one uncorrected bit in left/right stereo per hour. Large-scale integration error detection and correction Reed-Solomon and coding is used.

And just like in the RDAT, the rotary digital audio transport, so it is in the order of audio compact disk and the CD-ROM, that at any instant in time, no less than 64,000 bits of data are coming through the electronics, in transit, being corrected. This is doing the whole Reed-Solomon and coding business--you know, working out the syndrome, making the corrections, and all of that, doing those polynomial divisions.

And I would submit that there is another example: the audio compact disk, CD-ROM, the RDAT, rotary digital audio transport, are three machines which could not exist without large-scale integration of silicon.

While I was still working at the University of California at San Diego, people would repeatedly ask me-- the Chancellor and people like that; he was a psychologist of some kind; he can be forgiven for asking a question like this--but the question was repeated by other people, too. It is: When will silicon memory displace magnetic recording?

And I think the answer is clear from history. It will never do it. What will happen, as time goes on, is that the silicon memories that people use will get larger and larger and larger. If they are a megabit now, they will be 10 megabits in 1995 and 100 megabits in the year 2000.

Meanwhile, the recorder itself will have continued to increase in capacity. So, if floppy disks hold a megabit now, 1 megabit floppy disks won't exist ten years from now. They will all be silicon devices. Floppy disks will have moved up to 100 megabits and so on.

So, from the point of view of a recording person like me, I regard silicon technology as just being an enabling technology; it makes things possible.

So, that has finished, I think, all my discussion of the various threats that were thought. I've talked about electron beam recording of two kinds, optical recording of three kinds, including a magneto-optic tape recorder. I've talked about bubbles; I've talked about semiconductor memories, all of which must have cost Ampex an enormous amount of money and for which there is basically nothing to show.

And basically in the whole recording industry, there is almost nothing to show. It really is true--that sick joke--about starting off with a large fortune.

I want to just finish by telling you just a quick snapshot, just four foils about what has happened in recording.

(Showing of viewgraphs)

DR. MALLINSON: The most important thing that has happened since Mr. Poniatoff's day, 1956, is that the density of magnetic particles has increased enormously.

And on the top left here is a 1956 Ampex tape; and this is a 1966 Ampex tape, a 1976 tape, and a 1986 tape. It is pure g-Ferric oxide, more pure g-Ferric oxide, Cobalt-doped g-Ferric oxide, and metal particles. And there is a 1 micron marker down here, but you hardly need to see that. It's perfectly evident to you that something very, very dramatic has happened over that 30 year period in magnetic recording.

The density of particles in this iron particle tape is just 1,000 times higher than it was in the 1956 tape. You might ask: Well, why couldn't Ampex just jump to that immediately? The answer is that it is not making small particles that's difficult; it is formulating them, mixing them in the binder system, getting them uniformly dispersed--that's the difficult part. It seems to take a long time.

When you have a recording medium like that--and metal particle tape in my view, was the tape of the 1980s-- just as metal evaporated tape will be the tape of the 1990s--with a tape like that with 1,000 times more particles per unit volume, you get--if you are talking signal-to-noise power ratios--just 1,000 times the signal- to-noise ratio.

That means, track-width for track-width, wavelength for wavelength, 1,000 times the signal-to-noise power ratio, 30 dB signal-to-noise ratio.

(Change of viewgraph)

DR. MALLINSON: Ways of using that. For instance, the D-2 machine, to show you that it really exists; some of you heard discussion today about D-1 and D- 2. But there are two digital recording standards at the moment.

The D-2, which this is, takes the composite signal--that's the mixture of all the red-blue-green, if you will--and digitizes the whole thing at once; and a D-2 machine like this records that around 148 megabits per second.

The D-1 machine separates out the components--the red/blue/green, if you will--and digitizes each separately. And consequently, it requires a higher data rate; the data rate is 216 megabits per second.

So, there are two digital video recorders being made and in production in the world today: the D-1, which including overhead and eight audio channels and all of that, is 250 megabits per second; and the D-2 is 150 in this country and 160-something in Europe.

It's a rack-wired machine. The slot at the bottom takes cassettes.

(Change of viewgraph)

DR. MALLINSON: And the cassettes are called, not unreasonably, the large, the medium, and the small. I've got the dimensions on them there.

Let's look at the large one. It's 421 millimeters by whatever it is. It records for 3.5 hours-- 210 minutes--soaking up in Europe at 164 megabits per second. At the end of that 3.5 hours, it will have recorded 2 times 10^{12} bits. So, there is 2 terabits there.

A cartridge like that is rather like two VHS cassettes put side by side. So, if Mr. Poniatoff was still around, I would tell him: Look, Alex, we really have got the TV end in a drawer now. Now, we have this drawer; and you can put in probably five or six of those. You can have close to 10^{14} bits.

It is also interesting since we have been talking about alternative technologies to realize that one of those cartridges that costs about \$500.00--I'm sorry; they are called cassettes, not cartridges--has the same capacity as 30 IBM 3380 mainframe disk systems. And the mainframe disk systems cost \$100,000 apiece.

It is also equivalent in capacity to 40 two-sided 14 inch optical disks, which shows what I was saying earlier on about Ampex deciding that optical disks didn't have any future for a really high data rate or really large data storage is true.

1956

500Å spheres -- CoFe_2O_4



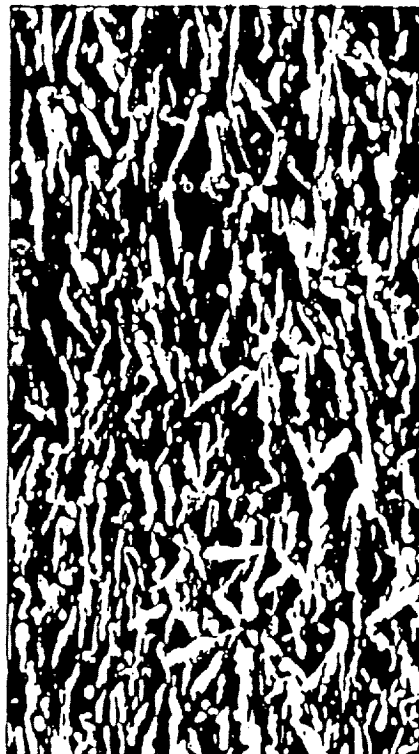
1966

5000Å needles -- $\gamma\text{-Fe}_2\text{O}_3$



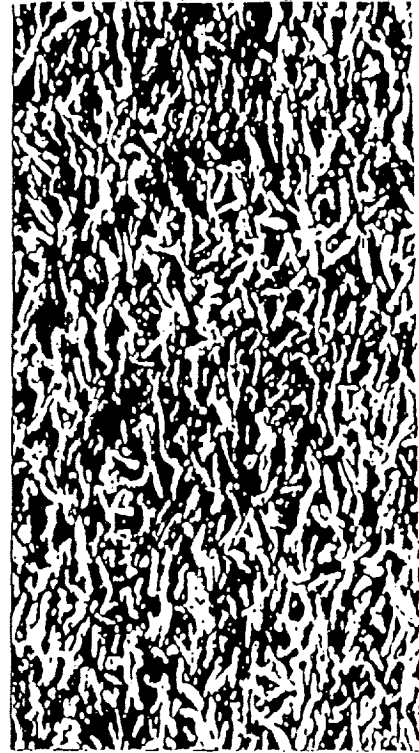
1976

2500Å -- $\text{Co-}\gamma\text{-Fe}_2\text{O}_3$



1986

1000Å -- Metallic Iron (MP)



10,000Å
1 μm

That D-2 is, in fact, I believe--and if anyone disagrees me, I would like to hear--it is the largest digital store, single module digital store, available in the world today; it is 2 terabits of data.

(Change of viewgraph)

DR. MALLINSON: And just to show that I am not terribly prejudiced, here is another development in rotary head machines, television recorders. This is Sony's experimental digital high-definition TV recorder.

The simplest way to think about high-definition television is to think that it requires about five times the number of bits per frame as a regular television. Regular TV means about 3 megabits per frame; HDTV is about 30 megabits per frame.

Another way of thinking about HDTV is that it is rather like high resolution computer graphics that you might find on some work station, except the difference is that in TV you've got to put up an image, 30 of them per second. And the data rate with a recorder like this has to record is 1,100 megabits per second, 1.1 gigabits per second. And it is achieved; it is a rotary head machine. There is a rotary head hidden under there.

In this particular machine, which was made in 1984--it's not new technology by any means--it was achieved with six heads. So, each one of the heads contacting the tape was running at something like 180 or 190 megabits per second.

When I left Ampex in 1984, it was considered conservative practice in the design of machines to run at 80 megabits per second. This is higher--190. If you ask me what's the world's record at the moment in the published literature for a magnetic head, actually writing and playing back, it is a mind-blowing 300 megabits per second. So, I think Ampex was right in its decision; there was no other technology that was going to achieve such high data rates as magnetic recording.

A recorder like this records HDTV at 1,100 megabits per second for one hour; and the capacity, when you've finished, of these 12 inch reels is--it says at the bottom--4.5 terabits, which is equivalent to 100 two-sided 14 inch optical disks.

The race is on in all the recording companies these days to try and reduce the number of heads. It would be nice to get the number of heads down to four, for instance, not six.

So, that's enough for the foils.

The conclusion I want you to draw is that I told you the story about 25 years of R&D in the Ampex Corporation and a great deal of money and time and very skilled people's efforts were spent on activities which led nowhere. They did not lead to the goal of having extremely large databases to be accessed at extremely high data rates.

And I think that is the end of my talk. Thank you for your attention.

(Applause)

DR. MALLINSON: Questions?

PARTICIPANT: Do you remember -- (Inaudible)

DR. MALLINSON: No.

PARTICIPANT: It was a holographic computer memory.

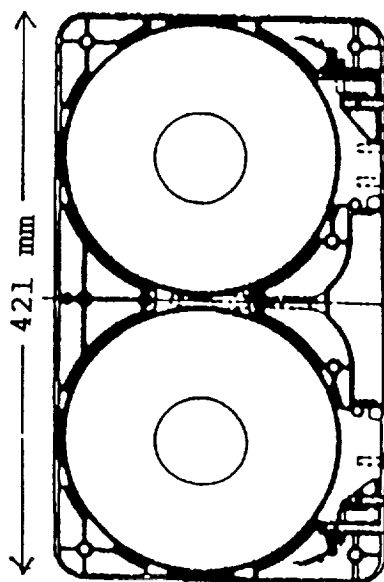
17" Wide (Rack Mount)

The front panel of the Ampex VPR-300 Digital Composite VTR includes a central control area with a digital display showing '00,21,19,24' and '0.86'. Below the display are buttons for 'ARM', 'FREEZE ON', 'FREEZE OFF', and 'DURATION'. To the left of the display are buttons for 'EDIT MODE', 'COLOR FRAME', and 'PLAY SPEED/FRAME'. To the right are buttons for 'FREEZE FIELD 1' and 'FREEZE DURATION'. The top section contains various input and output level meters, including 'INPUT VIDEO LEVEL', 'OUTPUT VIDEO LEVEL', 'INPUT AUDIO LEVEL', and 'OUTPUT AUDIO LEVEL'. The bottom section features a large array of buttons for various functions, including 'CUE', 'SEARCH', 'PREVIEW', 'EDIT', 'REVIEW', 'STOP', 'PLAY', 'PAUSE', and 'EJECT'. A large circular window is visible on the right side of the panel.

Ampex VPR-300
(D-2 Digital Composite VTR)

13 μ m THICK TAPE (MP)

210 mins/1622 meters

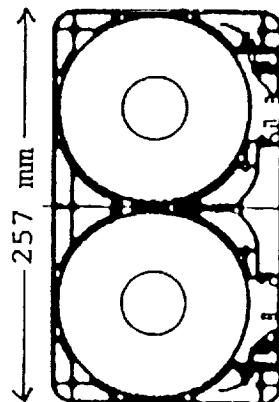


D2L

WORLD'S LARGEST CAPACITY

2 x 10¹² bits
30 x IBM 3380
40 x 14" optical disc

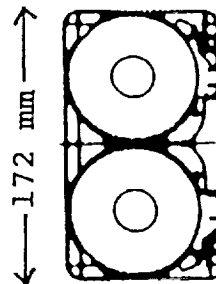
92 mins/708 meters



D2M

0.9 x 10¹² bits
13 x IBM 3380
17 x 14" optical disc

29 mins/222 meters



D2S

0.25 x 10¹² bits
2 x IBM 3380
6 x 14" optical disc

4.2.2 DIGITAL COMPONENT HDTV RECORDER

SONY HDD-1000 (Prototype only)



Sampling rates are: $74:27:27 \times 10^6$ sample/sec

Video bit rate is 1100 + o/h Mbs

8 digital audio (48/16) channels also recorded (2 Mbs/channel)

Gross bit rate: about 1200 Mbs

Playing time: 63 minutes

Capacity: 4.5×10^{12} bits (100 x 14" optical discs)

DR. MALLINSON: No.

PARTICIPANT: (Inaudible)

DR. MALLINSON: It was a scan?

PARTICIPANT: Yes. It was in **Time** magazine. (Inaudible)

DR. MALLINSON: No, no. I don't recall that.

PARTICIPANT: (Inaudible)

DR. MALLINSON: You know, that has to limit the time one talks. I could talk for hours about this; I find it fascinating. And I would like to be able to say a few things about what you are supposed to learn from all this. Well, I suppose the things that you can learn from this is-- and my remarks are just to do with tape recording--the goal of the Ampex Corporation is to record enormous databases, terabit databases, and access them at gigabit per second rates.

I think the message is that nothing will displace magnetic recording unless it is extremely simple; magnetic recording is done on a simple, chemically stable, featureless medium, it's cheap to make, it's cheap to implement.

I read all the excitement there is today about scanning tunneling microscopes; lo and behold, IBM can write IBM in some--do you remember what it was? Lithium or something? Xenon. It is evaporated away as soon as they let the temperature go up.

Then, the question is: Is there any reason to think that the very extreme high resolution of scanning tunneling microscopy--atomic resolution--will ever turn out to be a useful recording device? I think not.

Another one is atomic force microscopy, where you have a little magnetic tip that you measure the magnetic force over some substrate by vibrating it with a piezoelectric element and all of that. I think that one is even less likely to be a high density, fast access memory. In fact, I would add to the list of requirements that not only must the medium be cheap and featureless and stable, but there must be the requirement that you must be able to read every bit in at least 10 nanoseconds because 10 nanoseconds corresponds to 100 megabits per second.

And in fact, since all these magnetic recorders operate at that rate, and the 300 megabits per second I was talking about is only 3 nanoseconds to read each bit, a prerequisite really ought to be that you can read everything in just a few nanoseconds.

And then, there is the question of access time. Is there a way of accessing large distances? And is there a way of increasing the size of the recording medium to almost indefinitely large areas?

Questions like that, in atomic tunneling microscopy and atomic force microscopy, that I have found wanting.

PARTICIPANT: (Inaudible)

DR. MALLINSON: What I'm hearing is set size and tape height.

PARTICIPANT: Set size and tape height -- (Inaudible)

DR. MALLINSON: Cassette size?

PARTICIPANT: Yes. Real size -- (Inaudible)

PARTICIPANT: What's the difference in tension on the various sizes of cartridges that you displayed? And how does that affect -- (inaudible)

DR. MALLINSON: I hardly know how to answer. I mean, a standard tape is about a pound per inch of width, others are half-inch; and it is expected in video tape recorders where the tape contacts the head that the head shall be wearing down at the rate of about 10 microns per hour.

So, it's expected that the wear-out time of the head will be about 2,000 hours.

And a remark about cassettes may interest you. There is not a single cassette made--the D-1, the D-2, the 8 millimeter, the VHS, the late lamented Betamax--that even achieves 20 percent volume packing efficiency of the enclosure.

So, in order to have that convenience in magnetic recording in the cassette, which is simply the convenience that you can stick the thing in and pull it out at any time, you have given up a factor of 5 in volume packing density. 80 percent of the space inside the cassette is space, and they are still by far the highest volumetric packing factor devices in the world. With regard to another part of your question, I forgot to mention that it's funny that we used to be so worried about the tape contacting the head because any video tape now--even going by a VHS tape--to meet the specifications for VHS tape, it must stand one hour of still framing, which is something like 50,000 sequential passes of the head on the same track, without any measurable change.

And many of them will run for 10 hours.

PARTICIPANT: I have a question about tradeoff between areal density and the number of read -- (Inaudible)

DR. MALLINSON: There is nothing specifically about ID-1. I would be very surprised if any tape showed any deterioration in 100 reads. I'd even be surprised if it didn't go to 1,000 reads.

PARTICIPANT: (Inaudible)

DR. MALLINSON: I'm missing one word--the projection on the particle size?

PARTICIPANT: Yes.

PARTICIPANT: Could you repeat the question, please?

DR. MALLINSON: I think the question is: Do I have any idea of projections of the particle size? With iron particles that are in maybe RDAT and D-2 at the moment, they are around 1,200 Angstroms long and about 200 Angstroms wide. And it's interesting about that 200 Angstroms, incidentally, because I said this morning that various physicists have come up with the idea that there must be a quantum mechanical limit to recording density.

And my usual stock way is telling you the way to think about it some more is to tell them that in a single iron particle, 280 Angstroms in diameter, the flux flow is precisely one fluxon. Fluxons -- quantized magnetism has something to do with superconducting systems, not regular systems.

So, they are 1,200 by 200, and they will get smaller no doubt in the contact duplicating versions of DAT tape. The particle size is 800 angstroms long.

In the metallic thin films, the stuff I was talking about this morning, the metallic grains are 2 microinches by 2 microinches, 500 angstroms by 500.

There are media around at the moment which have a factor of 10 higher packing density than the iron particle tape. That's the name of the game with media; that's all you're doing. Raising the coercive force is a pain in the neck. It means more current has to go into the head; the medium must saturate; and who wants four times or five times the coercive force and 25 times the recording power?

The only reason you need it is to use smaller particles and get higher signal-to-noise ratios. Any more questions?

(No response)

DR. MALLINSON: Thank you.

REPORT DOCUMENTATION PAGE			Form Approved OMB No. 0704-0188	
<small>Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503.</small>				
1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE September 1992	3. REPORT TYPE AND DATES COVERED Conference Publication, July 23-25, 1991		
4. TITLE AND SUBTITLE NSSDC Conference on Mass Storage Systems and Technologies for Space and Earth Science Applications <i>Volume I</i>		5. FUNDING NUMBERS JON-933-656-80-02-26		
6. AUTHOR(S) Ben Kobler, P. C. Hariharan, and L. G. Blasso, Editors				
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) NASA-Goddard Space Flight Center Greenbelt, Maryland 20771		8. PERFORMING ORGANIZATION REPORT NUMBER 92B00093 Code 933		
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) National Aeronautics and Space Administration Washington, D.C. 20546-0001		10. SPONSORING/MONITORING AGENCY REPORT NUMBER NASA CP-3165, Vol. I		
11. SUPPLEMENTARY NOTES Kobler: Goddard Space Flight Center, Greenbelt, MD; Hariharan and Blasso: STX Corporation*, Lanham, MD. *Hughes STX Corporation as of October 1, 1991				
12a. DISTRIBUTION/AVAILABILITY STATEMENT Unclassified - Unlimited Subject Category 82		12b. DISTRIBUTION CODE		
13. ABSTRACT (Maximum 200 words) This report contains copies of nearly all of the technical papers and viewgraphs presented at the NSSDC Conference on Mass Storage Systems and Technologies for Space and Earth Science Applications. This conference served as a broad forum for the discussion of a number of important issues in the field of mass storage systems. Topics include magnetic disk and tape technologies, optical disk and tape, software storage and file management systems, and experiences with the use of a large, distributed storage system. The technical presentations describe, among other things, integrated mass storage systems that are expected to be available commercially. Also included is a series of presentations from Federal Government organizations and research institutions covering their mass storage requirements for the 1990s.				
14. SUBJECT TERMS Magnetic tape, magnetic disk, optical disk, mass storage, software storage		15. NUMBER OF PAGES 212		
		16. PRICE CODE A10		
17. SECURITY CLASSIFICATION OF REPORT Unclassified	18. SECURITY CLASSIFICATION OF THIS PAGE Unclassified	19. SECURITY CLASSIFICATION OF ABSTRACT Unclassified	20. LIMITATION OF ABSTRACT UL	