Current Trends in Data Storage Backup and Restoration

February 13, 2003

Tom Coughlin

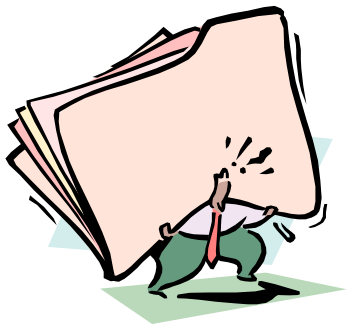
Coughlin Associates

www.tomcoughlin.com

Outline

- Storage Demand Drivers
- Backup and Recovery Trends
- Major Trends in Backup
  - Storage Hierarchy and Data Lifecycle
  - Tape Storage
  - Enhanced Backup
  - Disk Drive for Backup/Recovery
  - Form Factor Changes
  - Electrical Interface Development

Digital Storage Profile



Digital Exabytes Installed

Tape and removable storage
4x-15x all data on disk
Backup data from disk

New digital applications
Web apps
E-finance
Scientific
Education
E-medicine
E-mail/Video mail
Digital security
Entertainment video

Traditional applications

Linux	2.6%
Win2K	38.4%
Unix	47.9%
OS/400	0.2%
Other OS	5.3%
OS/390, z/OS	5.6%

	2000	2001	2002	2003	2004	2005
Primary Occurance	1.4	2.4	4.0	6.9	11.7	19.9
All Digital Data	5.6	11.9	20.2	34.4	58.5	99.5

in Exabytes

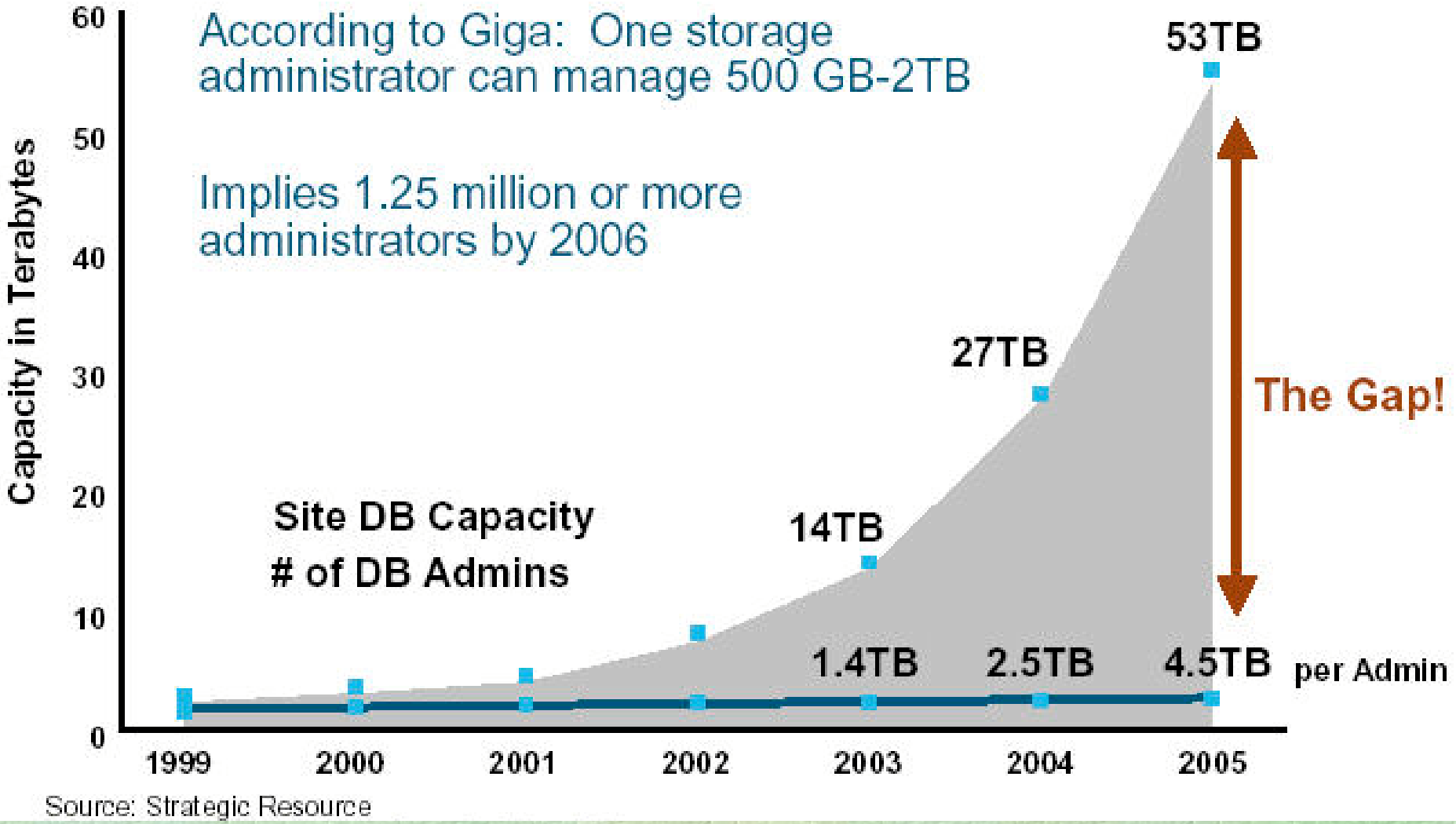
Source: Horison Information Strategies, University of California at Berkeley

Information Details

- Roughly 8 EB of digital data produced in 2002.
- 90% of data on disk is never or seldom accessed after 90 days+
- 90% of digital data is on removable storage*
- 80% of digital data is replicated data*
- Disk utilization is often as low at 35-45% ^
- Disk storage is the most expensive component in the data center

+Horison Information Services
*UC Berkeley
^Gartner/Credit Suisse

Need for Storage Administration

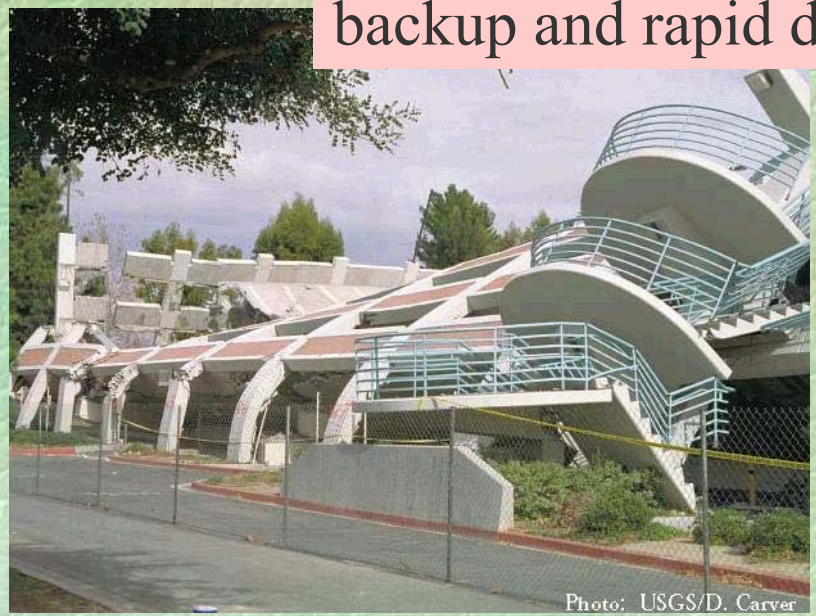


Source: Strategic Resource

Data Protection

- Provide Business Continuity Even If Data Is:
  - Accidentally Erased or Modified
  - Maliciously or Accidentally Modified
  - Corrupted
  - Catastrophically Lost
- Maintain an Accurate, Up-to-Date Copy of the Data
- Do Not Allow This Copy to Get Modified, Corrupted, or Lost
- Use This Copy to Get Back in Business Quickly

Disaster recovery Depends upon effective backup and rapid data recovery.



Costs of Site Downtime

Brokerage	$5.6M - $7.3M
Credit Card Authorization	$2.2M - $3.1M
Home Shopping	$87k - $140k
Airline Reservations	$67k - $112k
Subway Ticket Sales	$56k - $82k
Parcel Shipping	$24k - $32k
ATM	$12k - $17k

This is why rapid recovery is critical!

Gartner Group / Dataquest

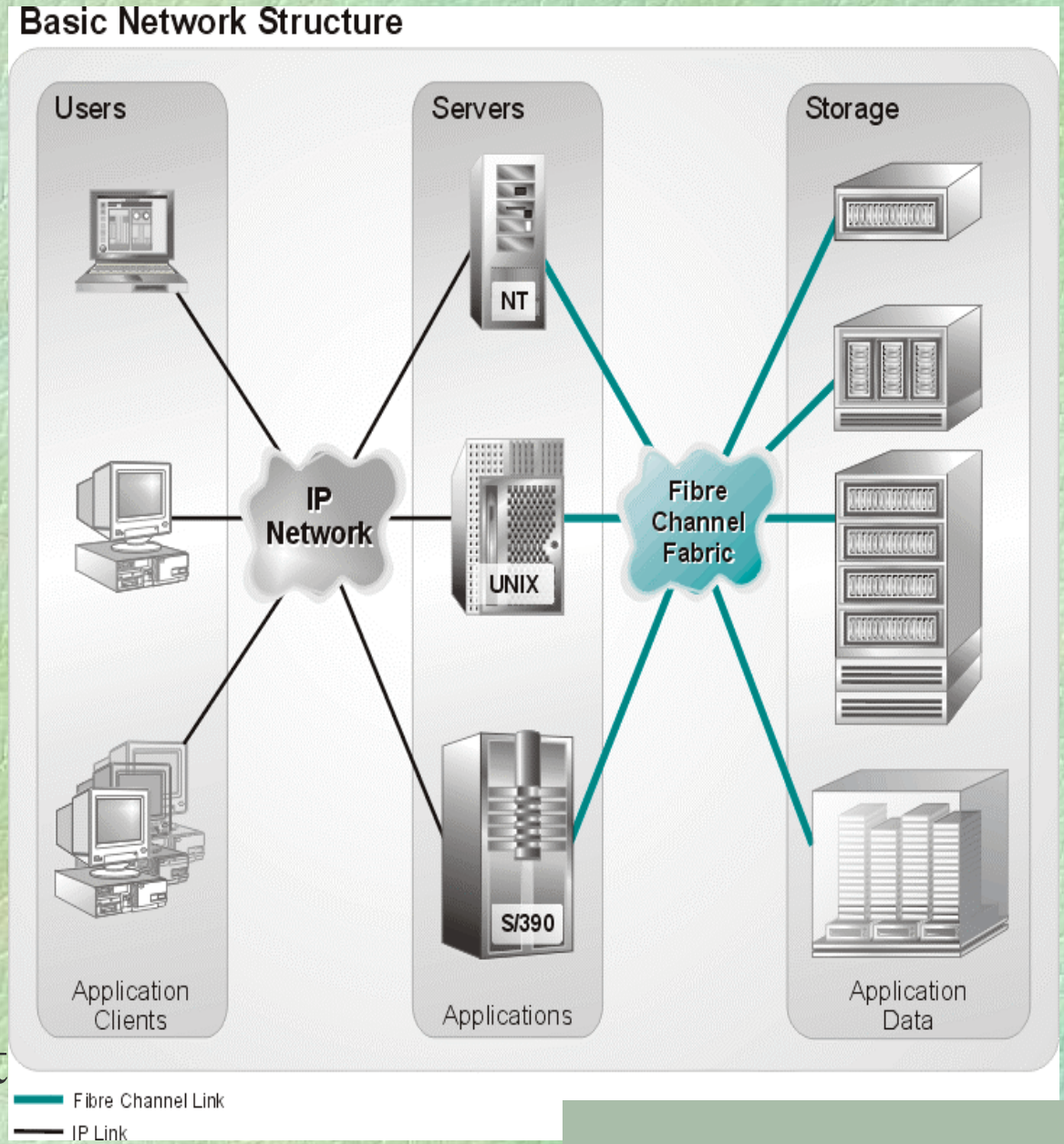
Many Backups are through Networks

SANs connect:

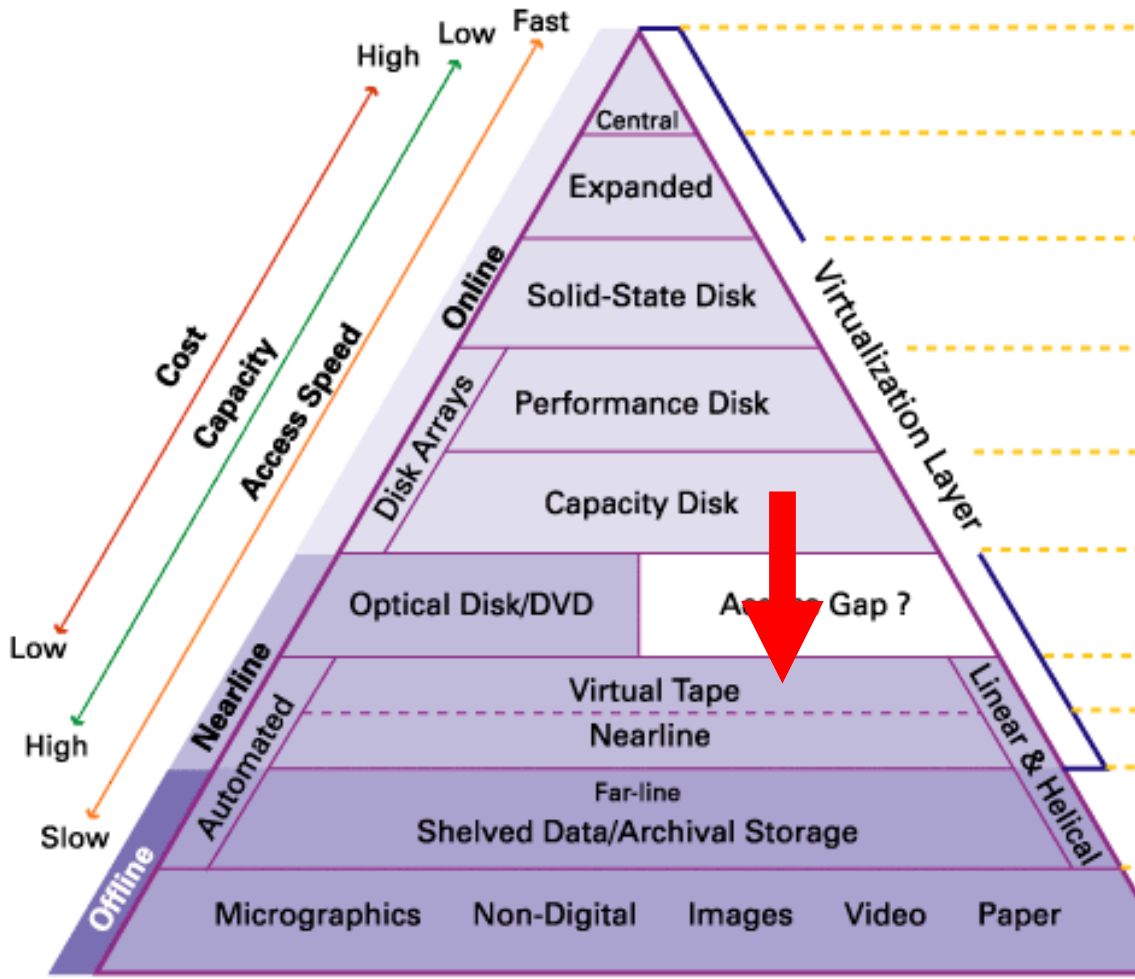
- Storage to Servers in the data center

IP connects

- Users to Servers on the LAN or Internet



The Storage Hierarchy



Positioning Parameters

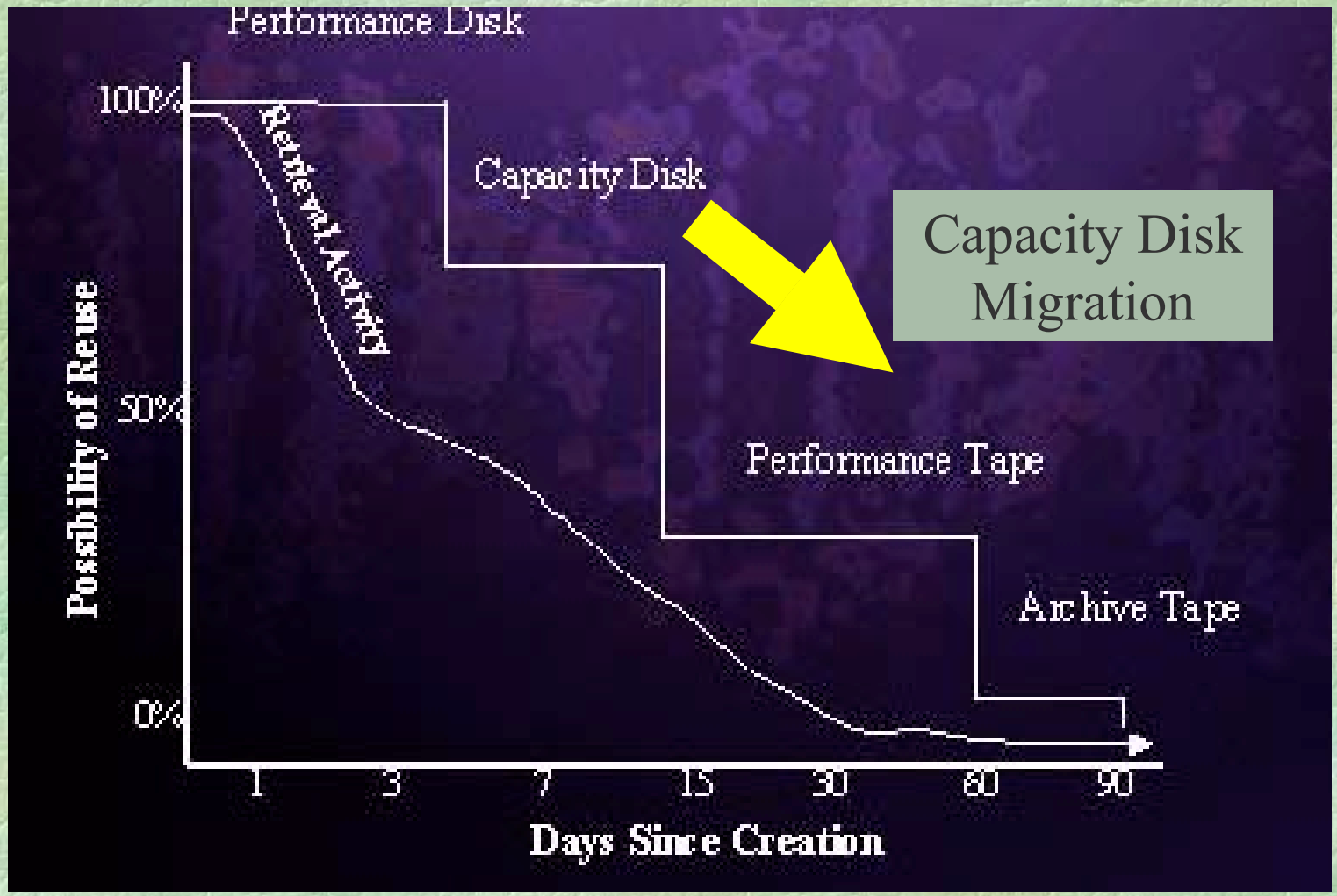
	Cost/MB	Capacity Limit	Access Time/ Performance
Central	$250-500	32 GB	24 GB/Sec
Expanded	<$250	16 GB	40 ns
Solid-State Disk	$10-15	25 GB	<.1 ms
Performance Disk	$.10-.15	N x GB/TB	4-10 ms
Capacity Disk	<$.10	NGB-40TB	10-25 ms
Optical Disk/DVD	$.10-.50	N x GB	5-10 secs
Virtual Tape	$.10-.30	TB-PB	5-10 secs
Nearline	<$.005*	TB-PB	5-10 secs
Far-line Shelved Data/Archival Storage	<$.001*	TB/PB	min, hr, day
Micrographics Non-Digital Images Video Paper	+$20	PB/EB	min, hrs, days

* Based on recording technology

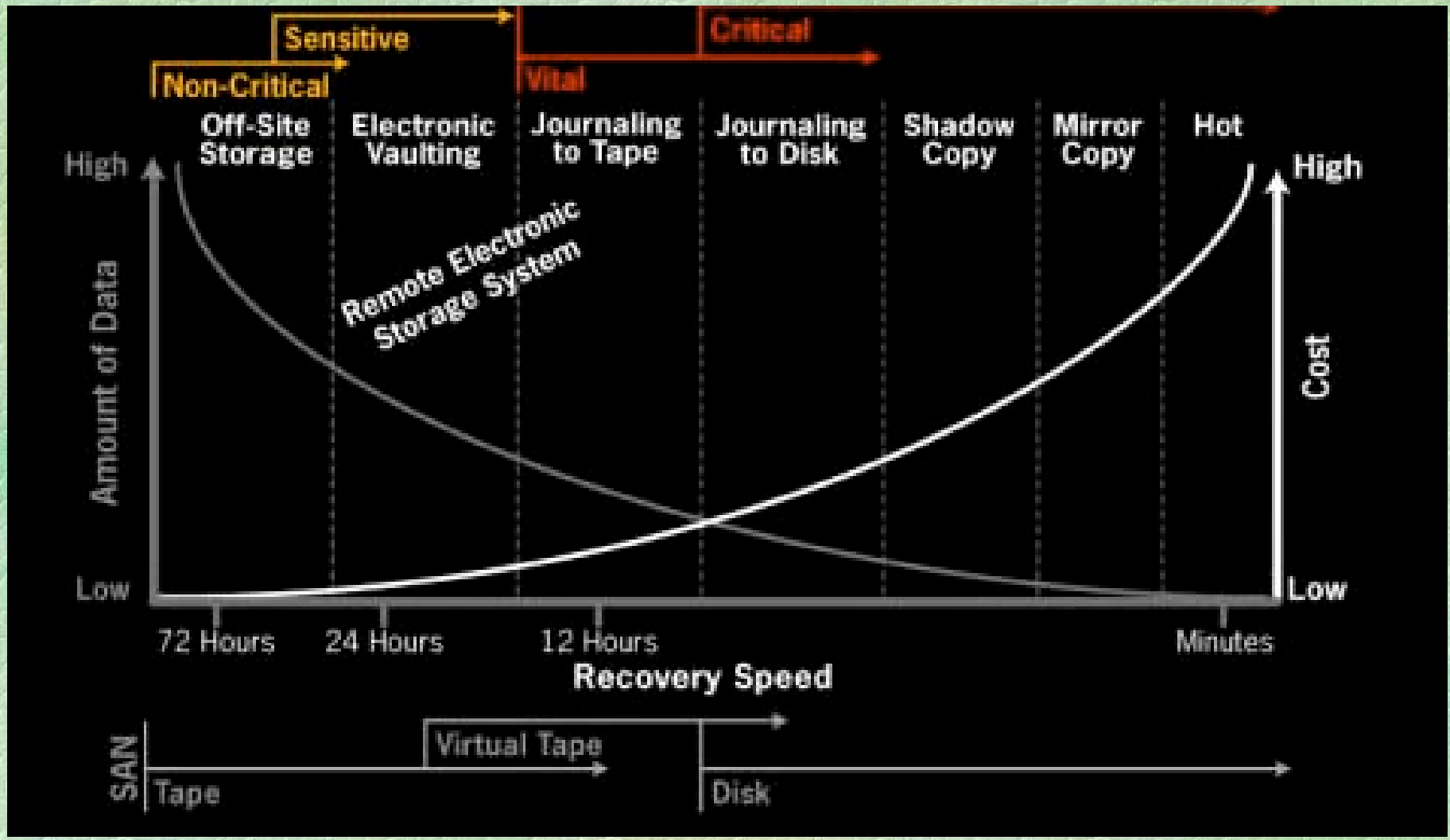
Source: *Horison*

www.horison.com

Data Lifecycle
(modified from StorageTek)



Recovery Time vs. Cost
(from StorageTek)



Tape Applications

- Largest single application is in back-up (>75%). Remainder is archive
- About half of average system price is for the autoloader systems and half is for the drives themselves
- Most backup using Veritas or Legato backup software, little NT or Unix.
- Biggest growth area is libraries for NAS or SAN systems

StorageTek Tape Library

Major Backup Tape Formats

AIT

LTO

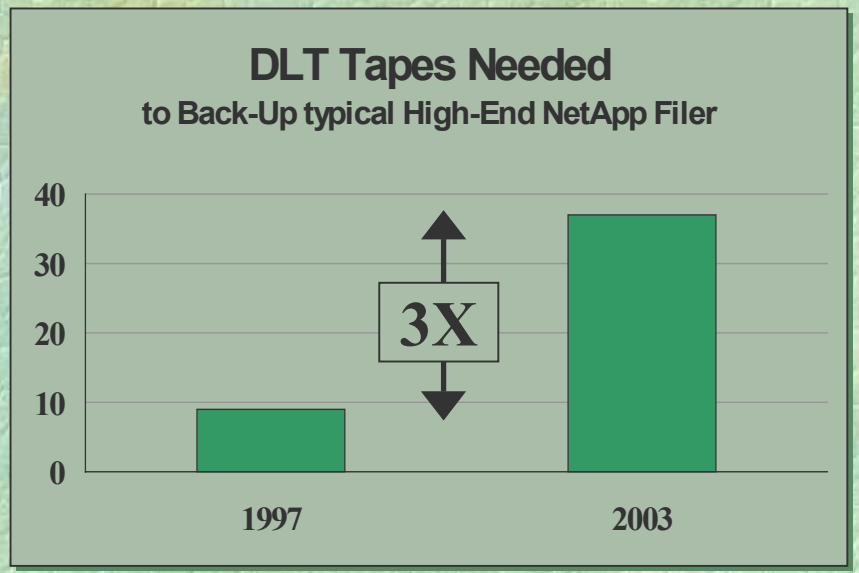
DLT

Tape Benefits

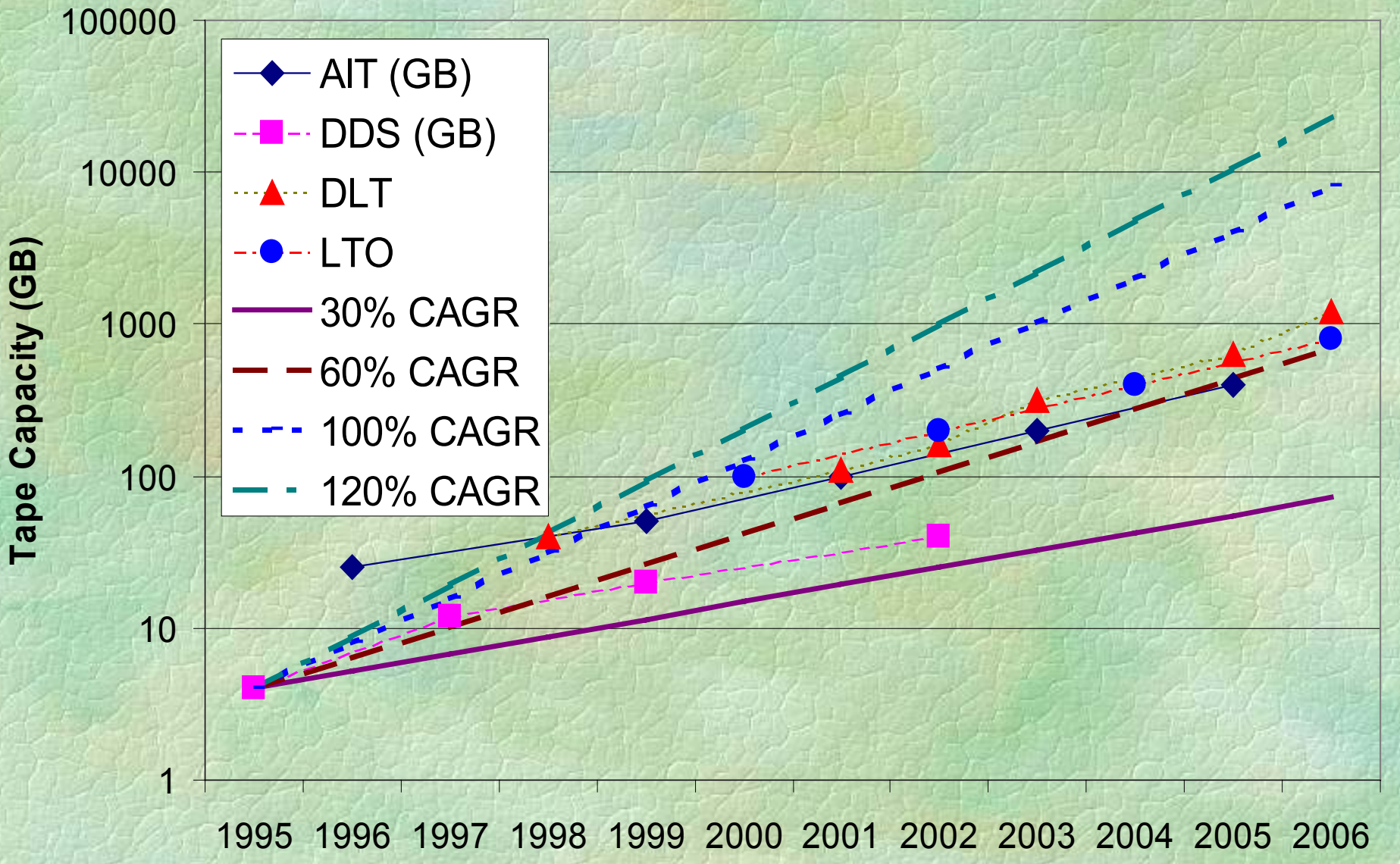
- Good Archival Medium
  - Shock Resistance
  - Packing Density
  - Transportability
- Cheap Media Cost

Tape Challenges

- Sequential Access
  - Slow data restoration
- Degradation During Long Term Storage
  - Re-tensioning, bleed through, …
- Lack of Scalability with Data Growth
  - Capacity
  - Throughput
- Periodic Verification Difficult
  - Especially if Offline



Tape Capacity Growth Trend vs. Technology



Tape Market Observations

- Tape prices tend to be very stable, <5% price erosion on systems per year
- Average drive price is about $5k (S-DLT)
- Average tape price is about $50 (S-DLT)
- Technology changes such as areal density growth and data rate improvements much slower than disk drives (<60% CAGR in Areal Density growth)

Enhanced Backup

- More than 80% of the cost of backup is operational costs, mostly manpower, to support backup.
- Since the core rate of tape technology development is different than disk backup, solutions with tape alone are scaling more slowly than the primary storage.
- This leads to a “backup crisis!”
- By enhancing traditional tape backup with disk based solutions we can help customers avoid a “backup crisis” and provide enhanced performance improvements as well.

EnhancedBackup.com

A NEW SPIN ON DATA PROTECTION

Search

EBSI
ENHANCED BACKUP SOLUTIONS INITIATIVE

SPONSOR BENEFITS
BECOME A SPONSOR
JOIN OUR USER GROUP
SIGN UP FOR NEWSLETTER

NETWORK STORAGE UNIVERSITY

Restores

Enroll in EBSI-Sponsored Seminar
Data volume keeps climbing. Backup windows keep shrinking. Efficient, cost-effective enhanced backup solutions can protect your company's data. Learn how from EBSI's founding members—the vendors driving this emerging storage scheme.
MORE

EBSI Puts Dynamic Leadership Team In Place
The Enhanced Backup Solutions Initiative (EBSI) has put together a strong leadership team and named an executive director to move forward with an aggressive agenda. Technology visionary Michael Peterson, has been named executive director. Founder of influential Strategic Research Corp., Peterson has proven expertise at incubating coalitions, including the Storage Network Industry Association (SNIA). Noted author and industry expert Jon William Toigo will serve as the board's end-user advocate. Industry leaders with more than a century of technology experience are backing EBSI's efforts.

MORE

TECH CORNER

VALUE PROPOSITION

- Enhanced Backup Helps Tame the E-Mail Beast
- Fighting the High Cost of Business Continuity
- The ABCs of Enhanced Backup

WHITE PAPERS

Driving for Dollars: ROI and Disk-Based Backup
The Giga Information Group determined the real-world economic impact of enhanced backup by studying six companies that had deployed or were about to deploy Network Appliance's NearStore R100. Giga also offers a financial model that can be used by any company that needs to run the numbers before deploying disk-based backup.
DOWNLOAD PDF 256K

Data At Risk: Can EBSI Help?
Despite vendor claims about faster robotic libraries, improved media capacity, enhanced mirroring, and the like, IT managers are struggling to develop, implement, and maintain effective data protection strategies. Meanwhile, a substantial percentage of mission-critical data remains at risk, and the potential for economic loss is staggering. What are the real roadblocks? Can EBSI's pledge to rise above the interests of any individual vendor help map the way out of this maze?

White paper available only to EBSI members.
MEMBER LOG-IN BECOME EBSI MEMBER

CASE STUDY

Enhanced Backup Restores Health to Hospital Storage
Hospitals operate at the speed of life. That makes slow, unwieldy storage solutions not just inefficient but potentially dangerous, as Leuven University Hospital knew all too well. One of Europe's largest medical centers, Leuven was saddled with an outdated direct-attached storage system supporting 3,000 PCs and 5,000 users on seven NT servers. Leuven wanted more than a mere upgrade. It wanted to streamline storage; simplify labor-intensive maintenance; roll out PACS (Picture Archiving and Communication System), a medical app that shares digitized X-rays and CAT scans across a network; add redundancy; and ensure scalability. It got what it needed, with some unexpected benefits, by deploying a Network Appliance™ Storage System and NearStore® Appliances.
READ CASE STUDY

MORE TECH CONTENT

PRODUCTS

Network Appliance NearStore
NearStore, from NetApp, delivers the access speed needed by near-online storage applications at a cost per MB between primary storage and tape/optical libraries. Data backups can run over the LAN or WAN..
MORE ON NEARSTORE

MORE PRODUCTS

READING STACK

- Veritas Says It's Beefed Up Backup Capabilities
- EMC Partnership Focuses on E-mail Archiving, SEC rules
- Serial ATA: It's Here
- Belluzzo Discusses Quantum Data Protection
- Separate SAN Fabrics for Disk and Tape?
- Banking On Low-Cost Data Mirroring
- Okapi Moves to Accelerate Data Backups
- The Lowdown on ATA Drives—and All the Rest
- iSCSI, Serial ATA Look Good to Cash-Strapped Customers
- Storage Virtualization: Watch the Forest, Not the Trees
- StorageTek Puts Backups on Autopilot
- EMC Spin-Off Targets Storage, Backup

MORE

PRESS ROOM

- Network Storage University Features Industry Visionaries
- Enhanced Backup Solutions Initiative Names Executive Director and Board
- EBSI and Network Storage University Team for Six-State Seminar Tour
- Quantum Rolls Out DX30 for General Release
- EBSI Takes Its Disk-Based Show On the Road: First Stop, N+I Atlanta

CMP Custom Media Solutions

EBSI
ENHANCED BACKUP SOLUTIONS INITIATIVE

IS SPONSORED BY

ATTO Technology, Inc.
AVAMAR
LEGATO
NetworkAppliance
qlogic Simplify
Quantum
TEK-TOOLS

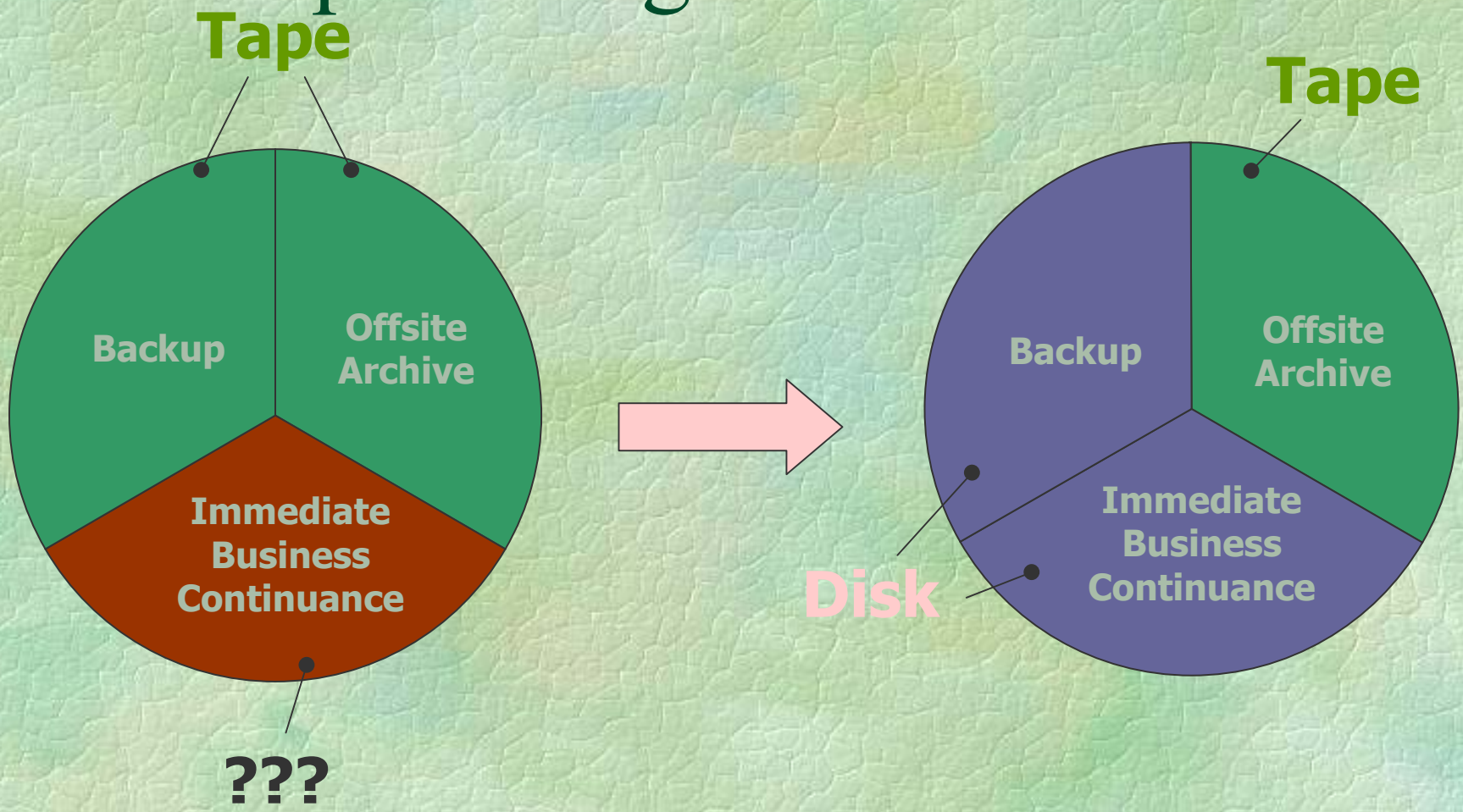
IN PARTNERSHIP WITH THE THEForsiteGROUP

Enhanced Backup

Exploit the Advantages of Disks to Protect Data

- Random Access
  - Fast Data Restoration
- Reliable
- Scalable
- Online Reliability Verification

Backup Paradigm Shift



Several Levels of Enhanced Backup

Level 3:
Continuous Backup with Read-Write Access

Level 2:
Changed-Block Backup with Read Access

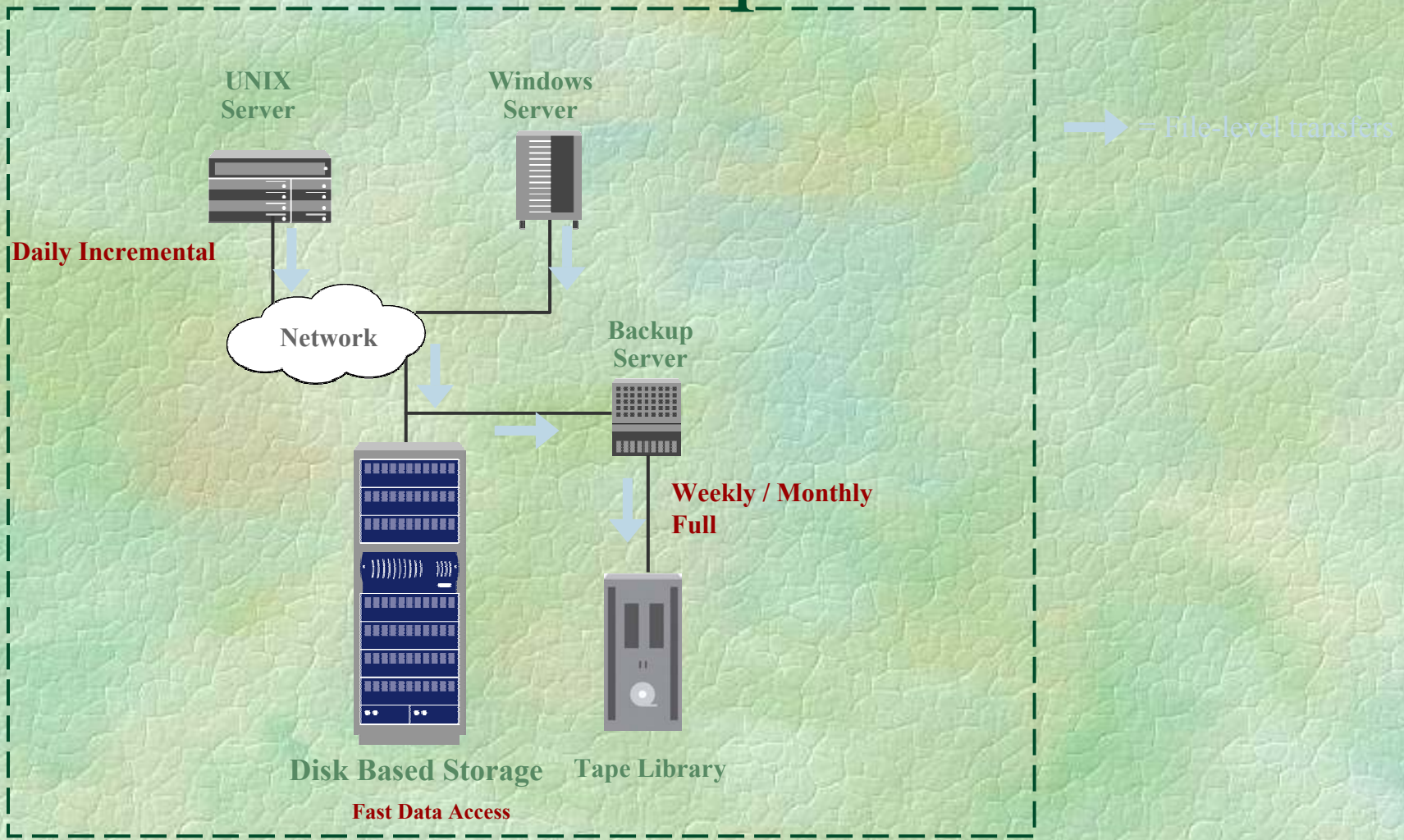
Level 1:
Backup to Disk as Tape Image

Enhanced Backup - Level 1

Backup to Disk as Tape Image

- Data on Primary Storage Is Backed up to Nearline Disk Storage Using Traditional Backup Software
- Data on Nearline Storage Is in Proprietary Format
- Nearline Storage Is Backed up to Tape for Archiving

Enhanced Backup - Level 1



Enhanced Backup - Level 1

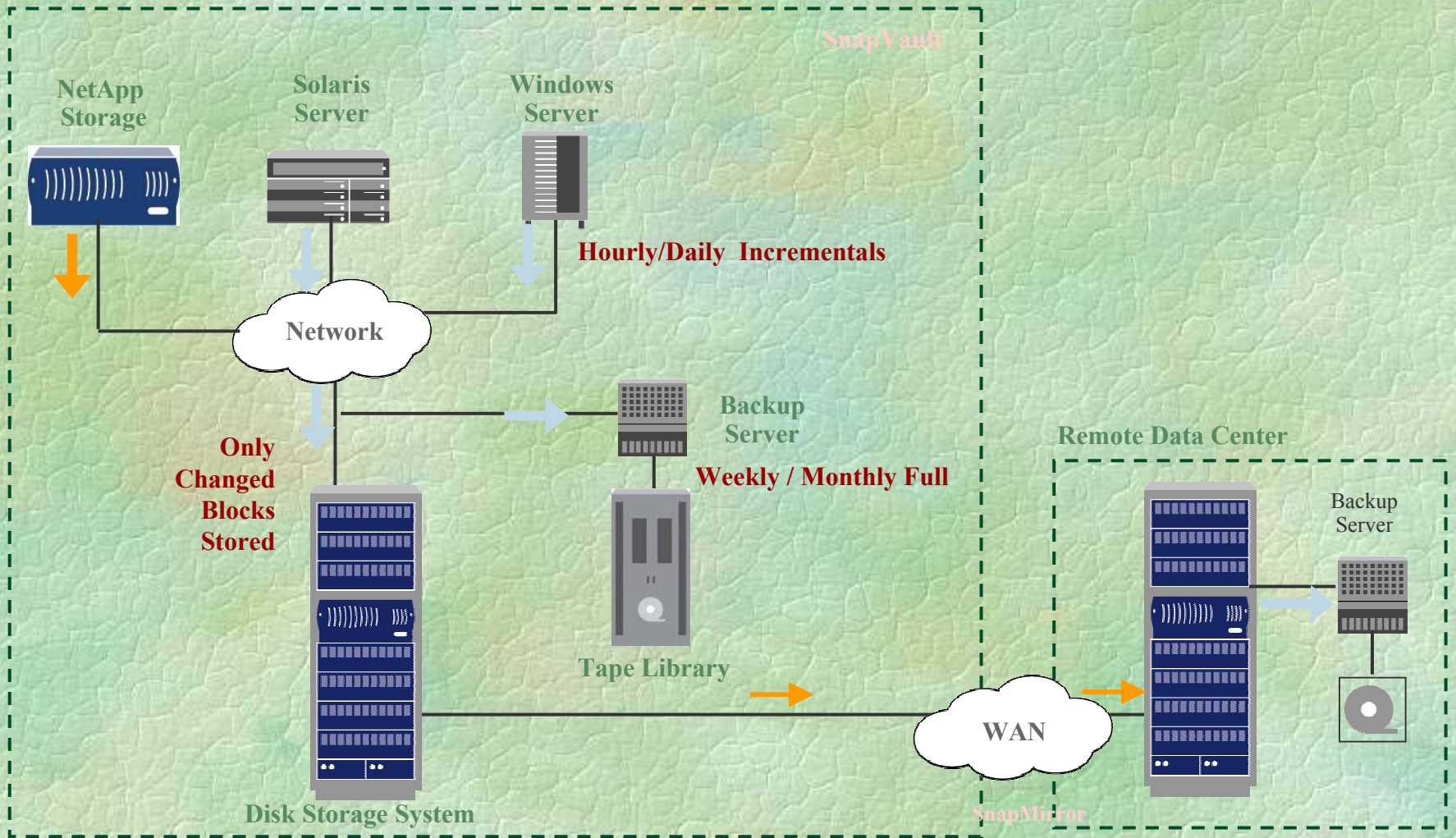
- Benefits
  - Faster Restores From Random-access Disk Storage
  - Eliminates the Need for Daily Incremental Backups to Tape
  - Integrates Into Your Existing Infrastructure
- Challenges
  - Lots of Disk is Required for Full and Incremental Backups
    - One Byte Changed Causes Entire File to be Backed up
  - Restore Process Still Requires Human Intervention
    - Backup Copy Cannot Be Directly Accessed
  - Backing up Remote Offices Is Not Practical Using This Approach
    - Requires a Robust WAN Network

Enhanced Backup - Level 2

Changed-Block Backup with Read Access

- **Data Is Backed up to Nearline Disk Storage**
  - Only the Initial Backup to Nearline Storage Is a Full Backup
  - All Subsequent Backups Transfer Changed Data Only
    - Only Changed Blocks Are Stored
- **Backup Data on Nearline Storage Is in File Format**
  - Can Be Browsed By Users

Enhanced Backup (Level 2)



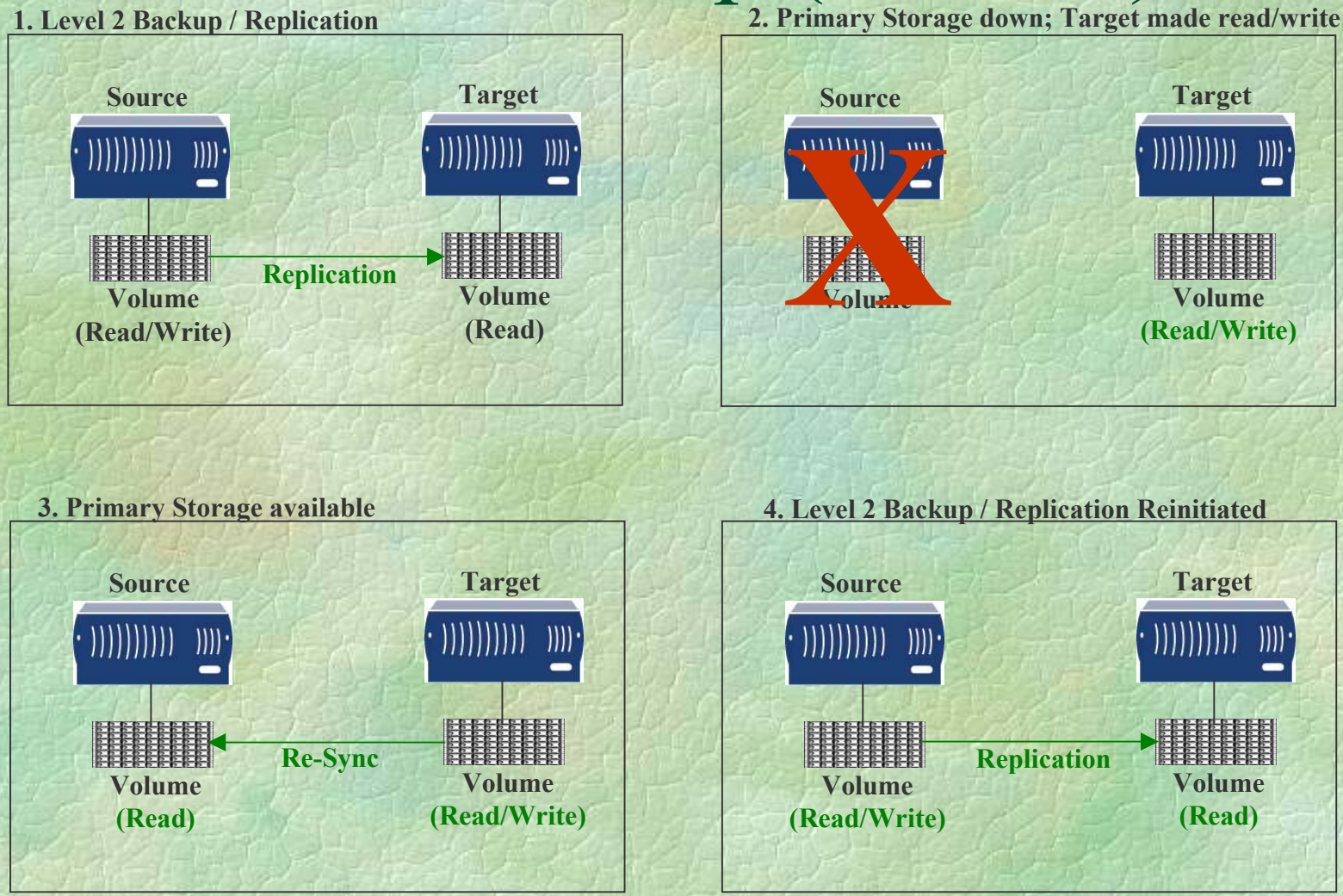
Enhanced Backup (Level 2)

- Benefits
  - Superior Data Protection
    - More frequent backups can be done and kept online
    - Immediate verification of backup data
  - Fast Backups and Restores
    - Shrinks/eliminates the backup window
  - Lower Backup Infrastructure costs
    - Less storage utilized to store backup copies
    - User initiated file restores
- Challenges
  - Files Need to Be Restored Before Use
    - Restore Is Delayed Until a New System or Free Disk Space Can Be Located
  - Doesn’t Solve Immediate Business Continuance
    - Separate Solution Required

Enhanced Backup (Level 3)

- Continuous Backup with Read-Write Access
  - Backup Data on Nearline Storage Can Be Made Write-able in the Event of a Disaster
  - Once the Primary Storage Is Available, the Data on the Nearline Storage Can Be Re-synced With the Primary Storage

Enhanced Backup (Level 3)



Enhanced Backup (Level 3)

- Benefits
  - Superior Data Protection
    - More Frequent Backups Can Be Done and Kept Online
    - Immediate Verification of Backup Data
  - Lower Backup Infrastructure Costs
    - Less Storage Utilized to Store Backup Copies
    - User Initiated File Restores
  - Solves Backup and Business Continuance Issues
    - One Solution
- Challenges
  - New Paradigm

Addressing Traditional Backup Pain Points

Traditional Backup Pain Points	Backup to Tape	Level 1	Level 2	Level 3
Primary Storage impact during backup	X	●	✓	✓
Backup window shrinking is an issue	X	●	✓	✓
Restoring data takes a long time	X	●	✓	✓
Takes a long time to verify backup data	X	X	✓	✓
Backups consume a lot of tape media	X	●	✓	✓
Backups consume a lot of network bandwidth	X	X	✓	✓
Backup & restore process fails thereby requiring constant monitoring	X	●	✓	✓
Restores normally require administrator involvement	X	X	✓	✓
Remote backups are not dependable and costly to manage and administer	X	X	✓	✓

X Does not address
● Helps address
✓ Fully addresses

Nearline and Enterprise Drives

Seagate Cheetah Product
73.4 GB, 15,000 RPM, FC/SCSI

Western Digital Caviar Product
200 GB, 7,200 RPM, PATA

Maxtor MaxLine Product
320 GB, 5,400 RPM, SATA

Western Digital Raptor Product
36.7 GB, 10,000 RPM, SATA

ATA-Based Storage Systems

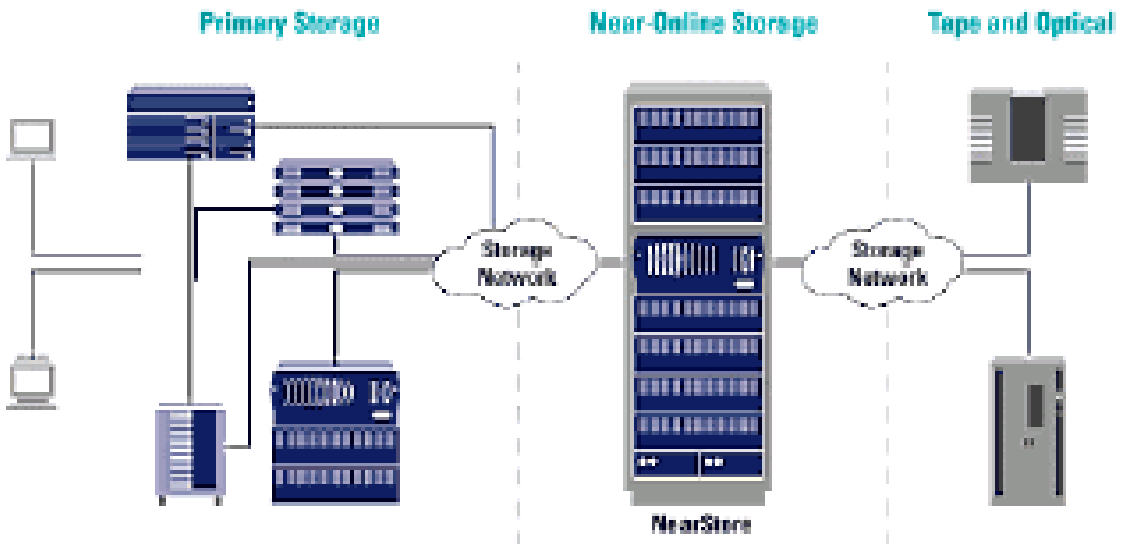
Quantum DX30
The DX30 separates backup functions from archive functions to optimize the data protection process.

Nexsan ATABeast Nexsan's 14 TB for 7 cents a MB

STK Bladestore product uses 5-3.5 inch drives on blade acting as one drive to a fibre channel output

Nearline Storage

- **NearLine storage is most often a rack of HDD storage shelves and is used for low latency backup (caching for tape archives)**
  - Sits between tape archives and primary storage arrays
  - Parallel ATA-based today (low cost drives are a must)
  - FC SAN connections
- **NearLine storage requires reasonable reliability and reasonable performance (cost is more important)**
  - Focus is on cost and capacity (cost/Mbyte)



Disk Drive Trends

- Increasing storage and lower $/GB
  - Currently 60 and 80 GB/3.5 inch disk
    - Maxtor 320 GB, 4 disk, 5400 RPM
    - Maxtor, WD 200+ GB 7200 RPM
  - Next year 120-160 GB/3.5 inch disk
  - Within 2-3 years 1 TB 4-disk drive will happen!
- New serial interfaces
  - Serial ATA (SATA)
  - Serial SCSI (SAS)
- Growing use of external drive boxes with USB or 1394 interfaces
- New small form factor drives for mobile devices
  - 1.8 inch 20+ GB drives and small drive developments

External Drives (USB or Firewire) or with small NAS devices on a LAN

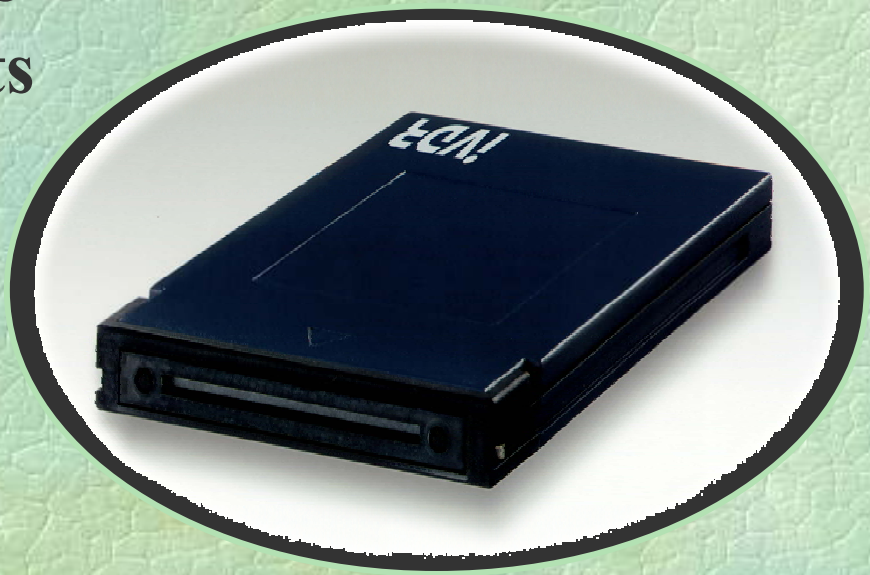
Maxtor PS5000 with one-touch backup

SNAP Storage Appliances

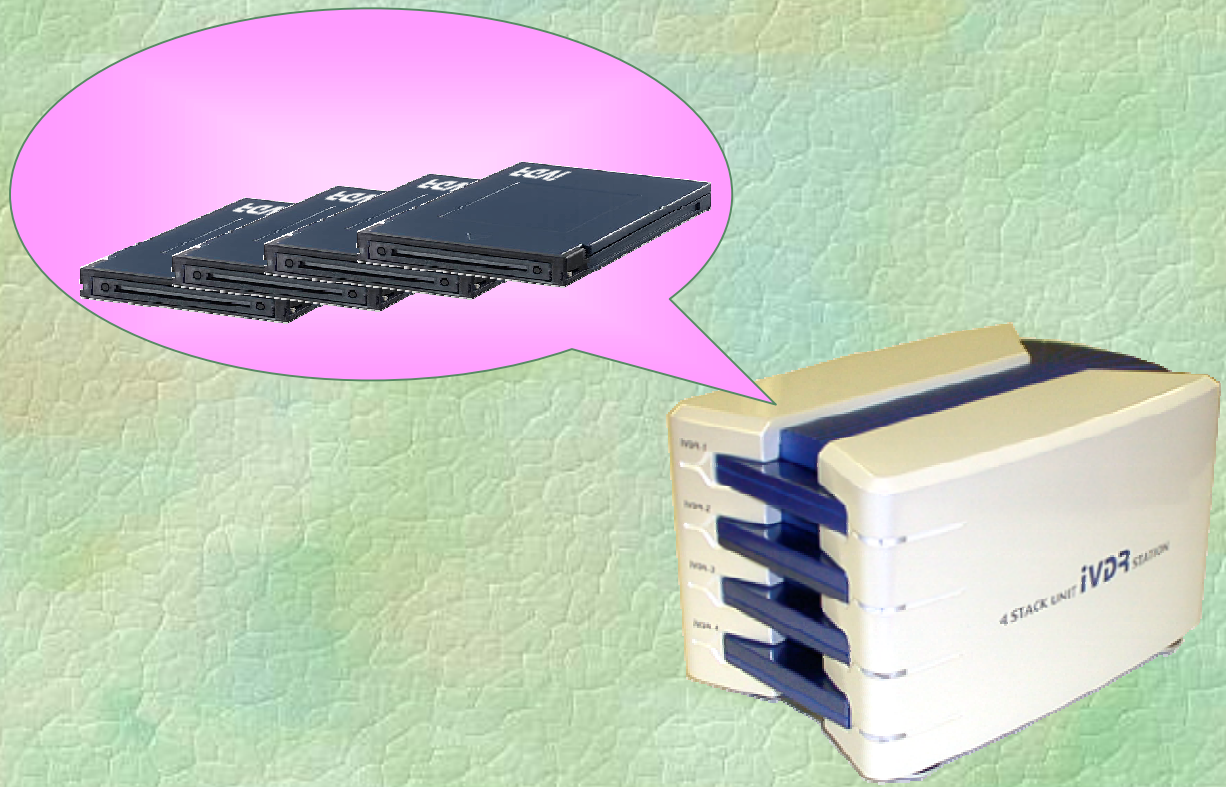
iVDR

Information Versatile Disk for Removable usage

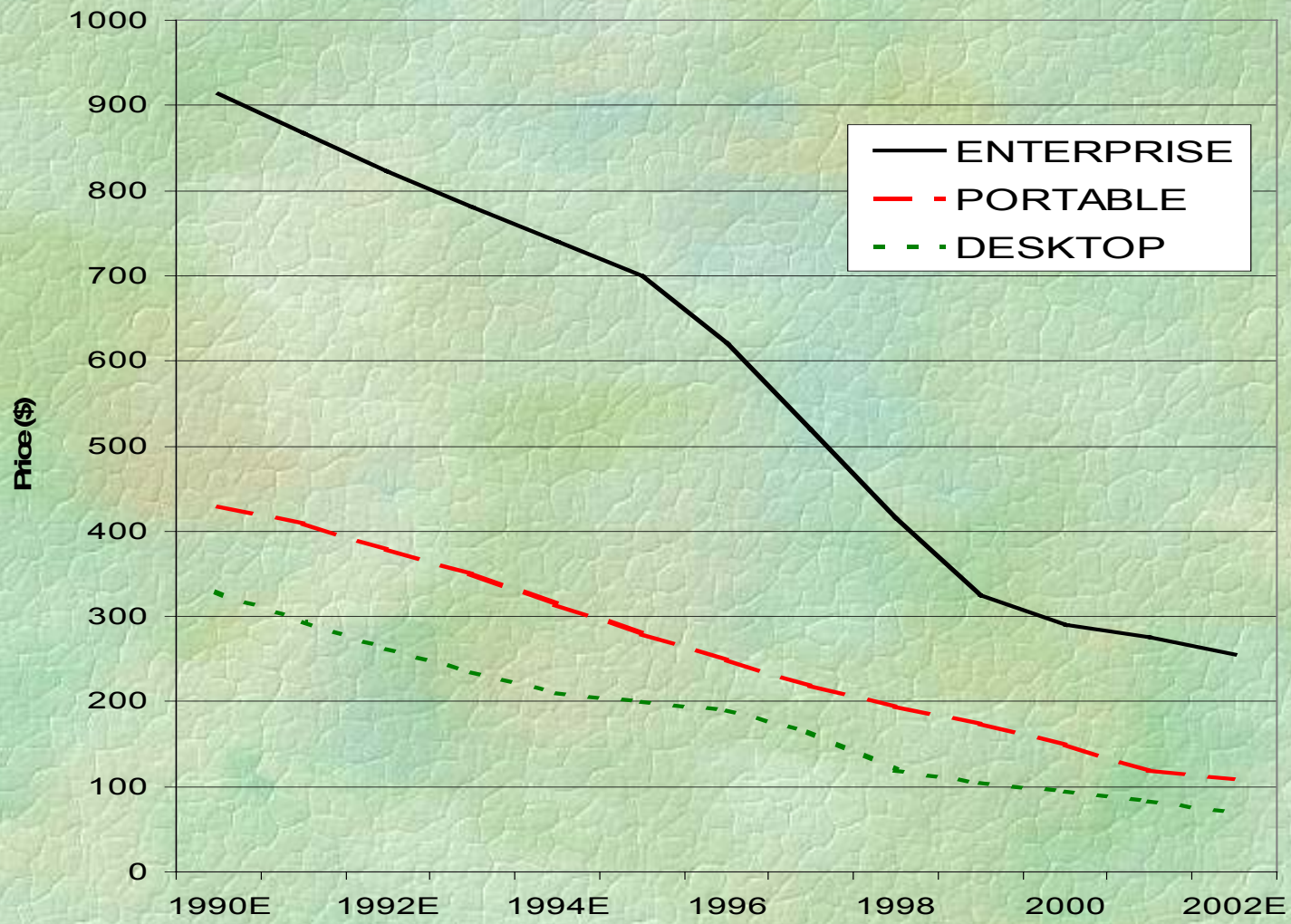
- **Common HDD platform for PC and Consumer AV usage**
  - **regardless of products and manufactures**
- **Compact and Removable**
- **Large Capacity and High-Speed Access**
- **Content/Data Protection**
- **Open Standard**



Possible Backup NAS Device using iVDR drives

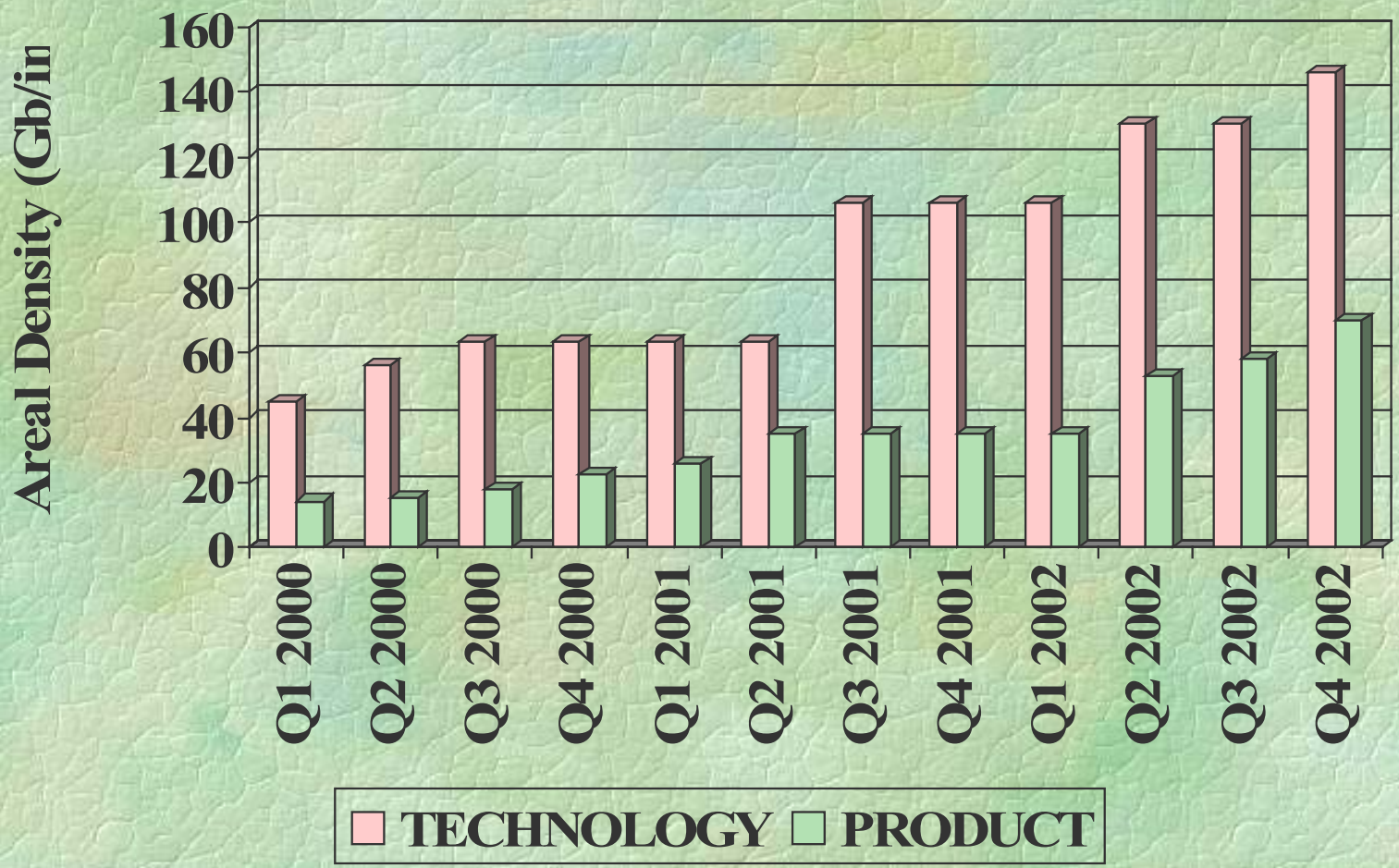


Estimated ASP Trends



AREAL DENSITY PROGRESSION

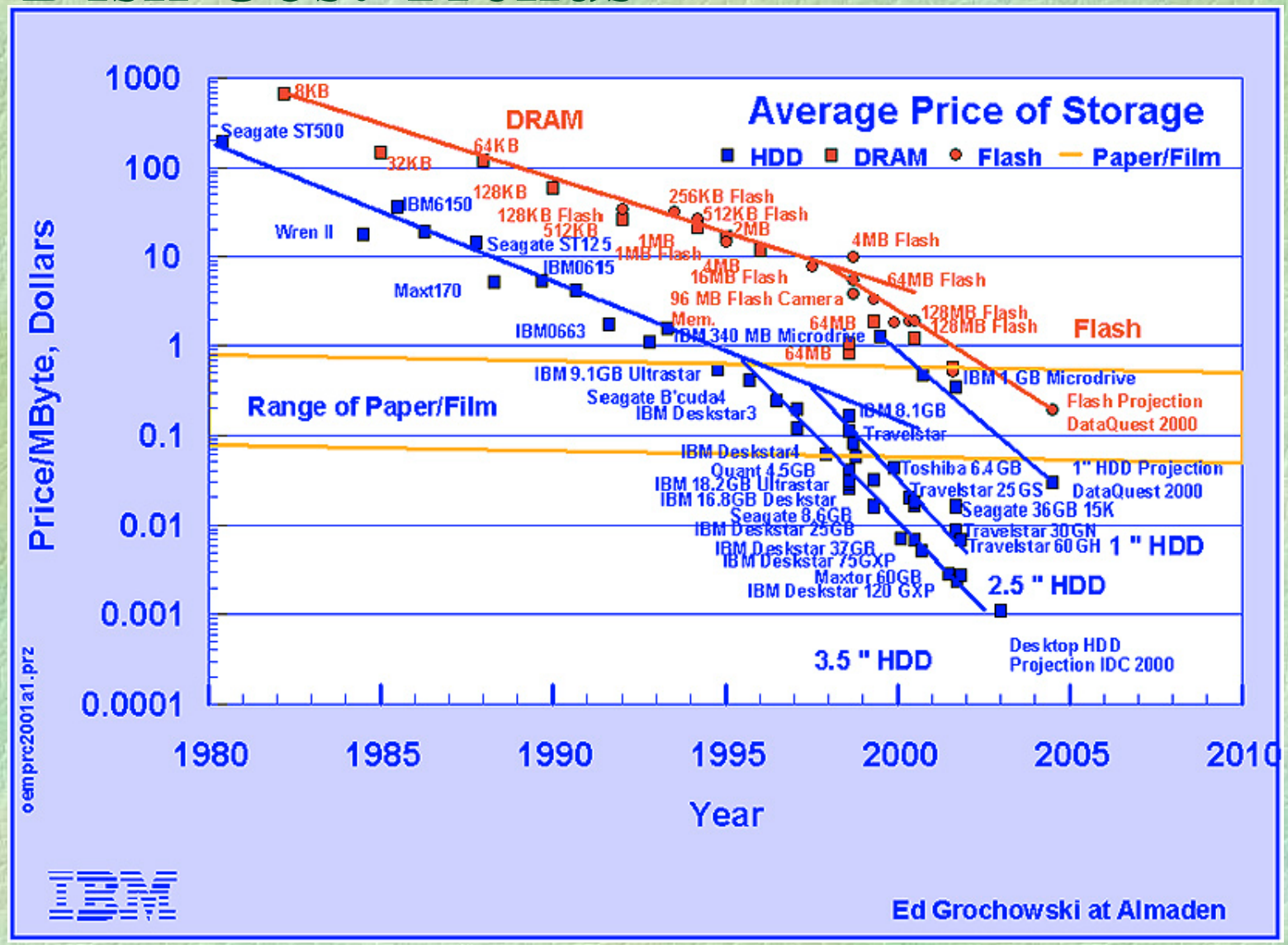
(Source: PRC, 2002)



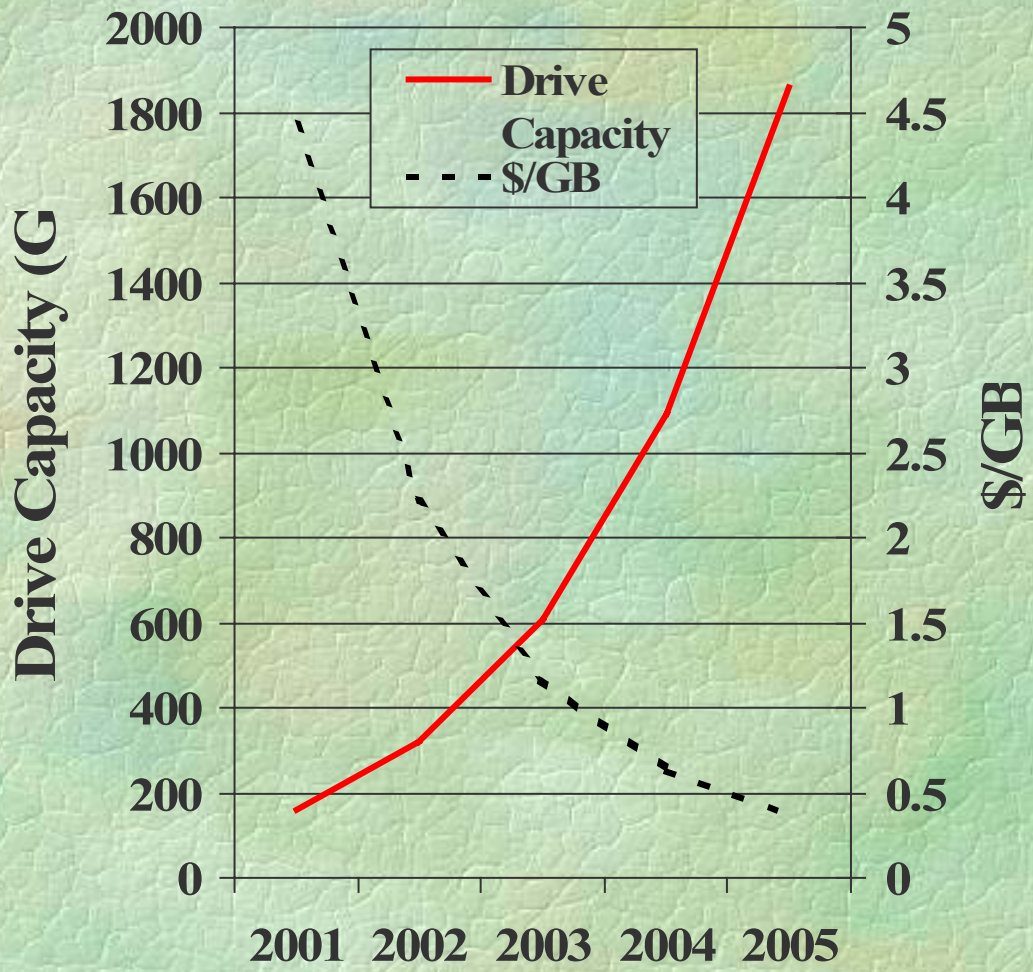
SHIPPING PRODUCT AREAL DENSITY PROJECTIONS

Year	Areal Density CAGR	95mm Avg. Capacity Per Platter
2000	120%	15
2001	100%	30
2002	90%	60
2003	80%	108
2004	70%	184
2005+	60%	294

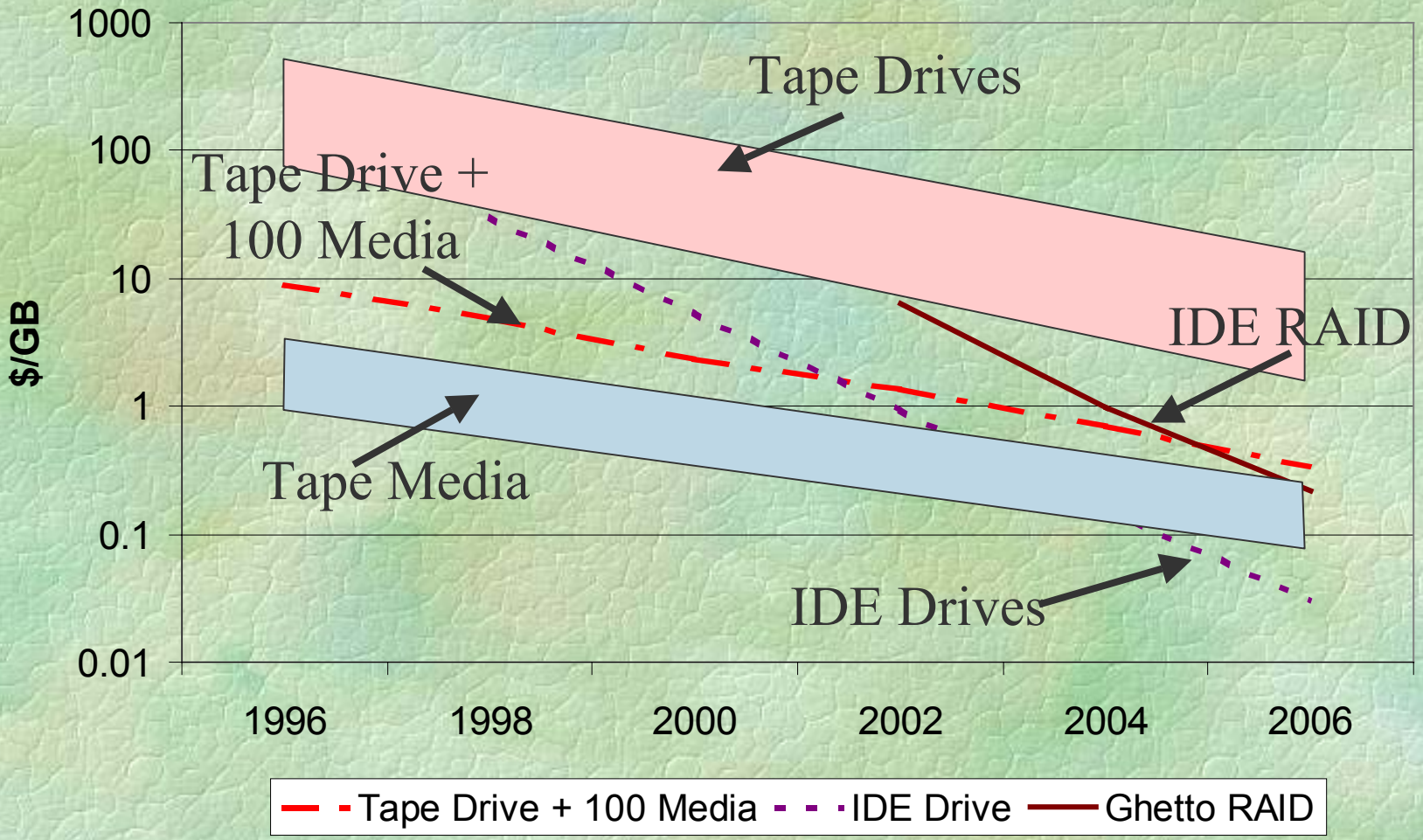
Disk Cost Trends



3.5 Inch ATA Network Storage Drive Capacity and Price/GB



As low cost disk drive storage decreases in price it offers greater economy to disk to disk backup and the use of disk drives for backup cache.



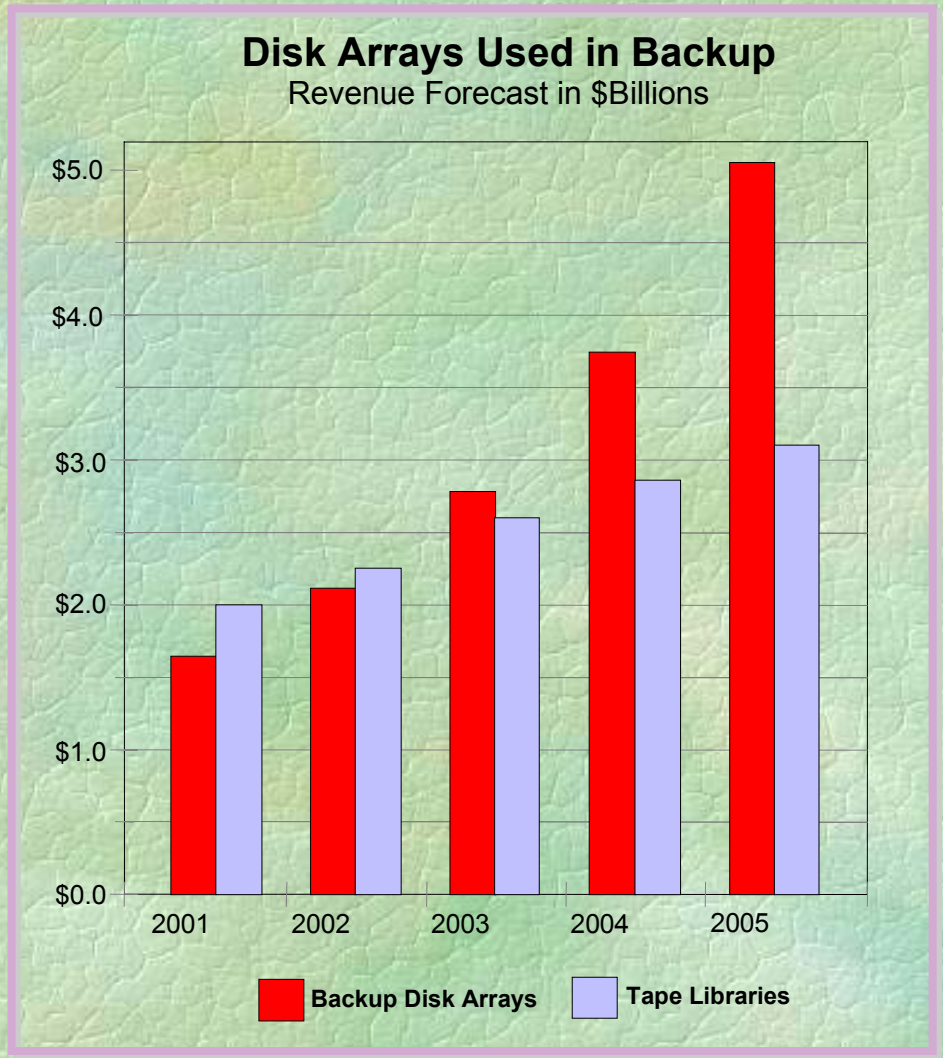
Comparison of Straw Man DLT Tape vs. IDE Disk Backup System

(Note that Tape has 2:1 Compressed capacity vs. disk drive native capacity)

Attribute	DLT Tape Libray	IDE Drive Drive
Access Time	60 sec	<15 ms (>4000 X faster)
Data Rate	6 MB/s	>46 MB/s (>7 X faster)
Removability	Yes (Cartridges)	Could be (drive carriers)
A. D. CAGR	<60%	>80%
	Sequential Access	Random Access

DATA PROTECTION MARKET OPPORTUNITY

- Backup Arrays include
  - **Virtual Tape, D2D Backup, Point-in-time Backup, Snapshot Backup**
- Backup Array revenue grows to $5.1B in 2005 offsetting the Tape Library Market
  - **Tape Library growth reaches $3.1B in 2005**
- Disk usage expands as a secondary data protection device relegating tape to an archive role
  - **Tape libraries are the central automated archive repository**
  - **60%+ of mainframe data is now protected by disk – Virtual Tape**



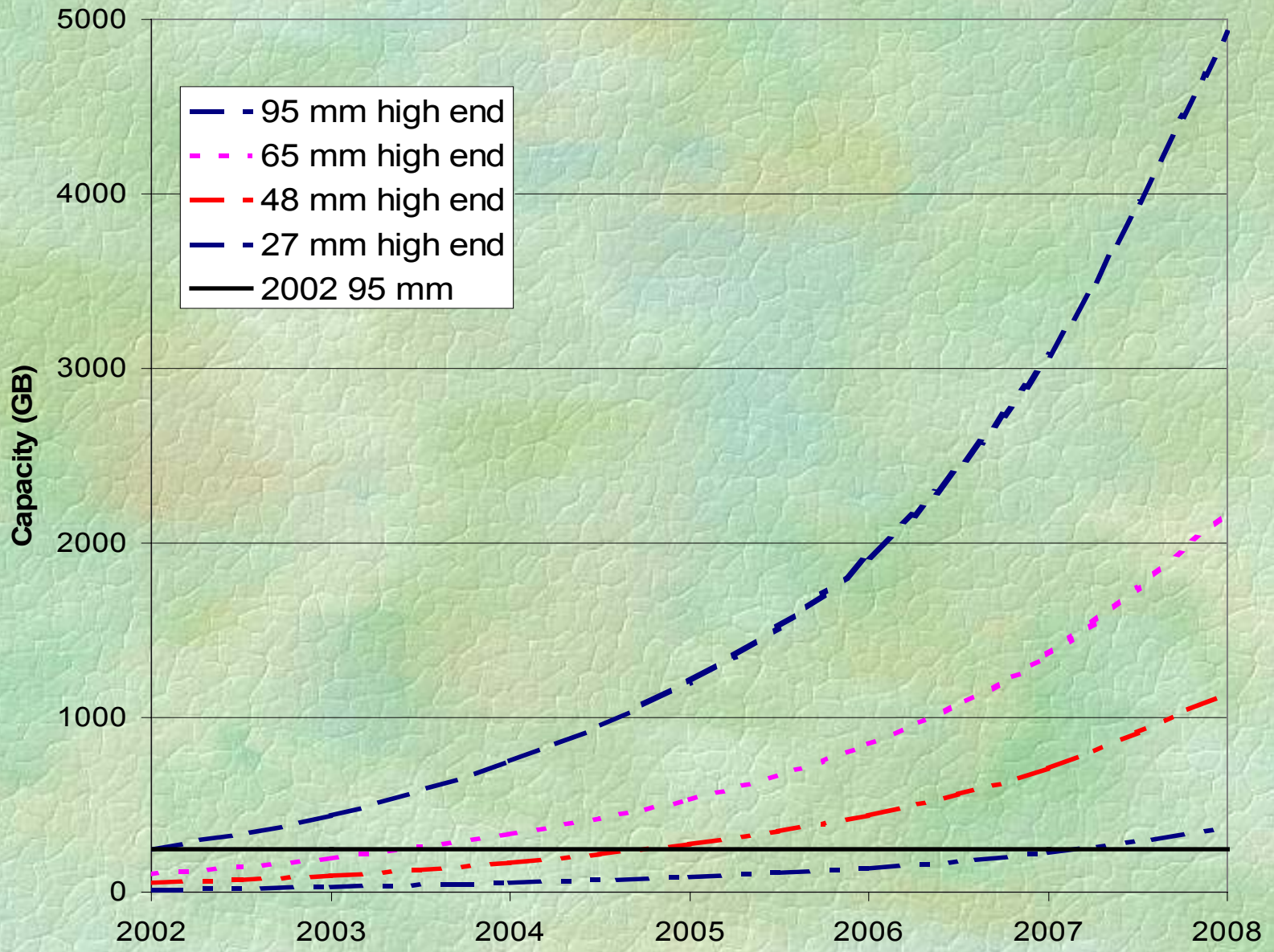
Strategic Research Corp., Nov. 2002

Transition to Smaller Form Factors

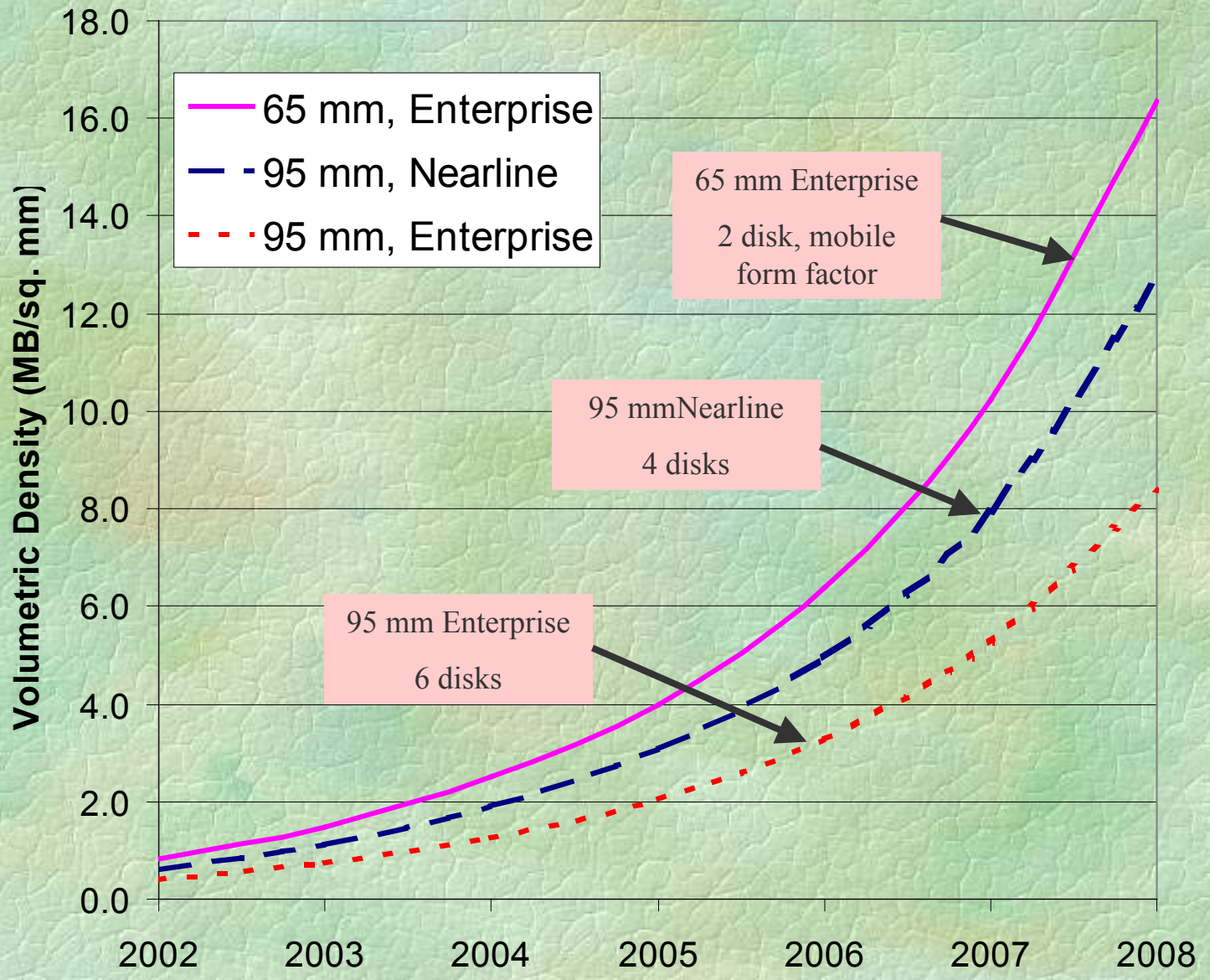
- 2.5 inch most popular mobile computer drive form factor.
- 1.8 inch mobile computers now appearing, smaller size drives???
- 60-65-mm disks used in 15k RPM enterprise disk drives (although not yet in 2.5 inch form factor box). Cooling issues
- For new consumer products size and volume will become important.
- Dense server and storage environments favor many more smaller drives. This also gives better performance since the time to data is faster for smaller form factors
- New consumer electronics initiatives using smaller form factor disk drives such as the Japaneses iVDR consortium.
- In volume 2.5 inch drives should be as inexpensive or less expensive per box compared to 3.5 inch disk drives.

Capacity vs. Form Factor

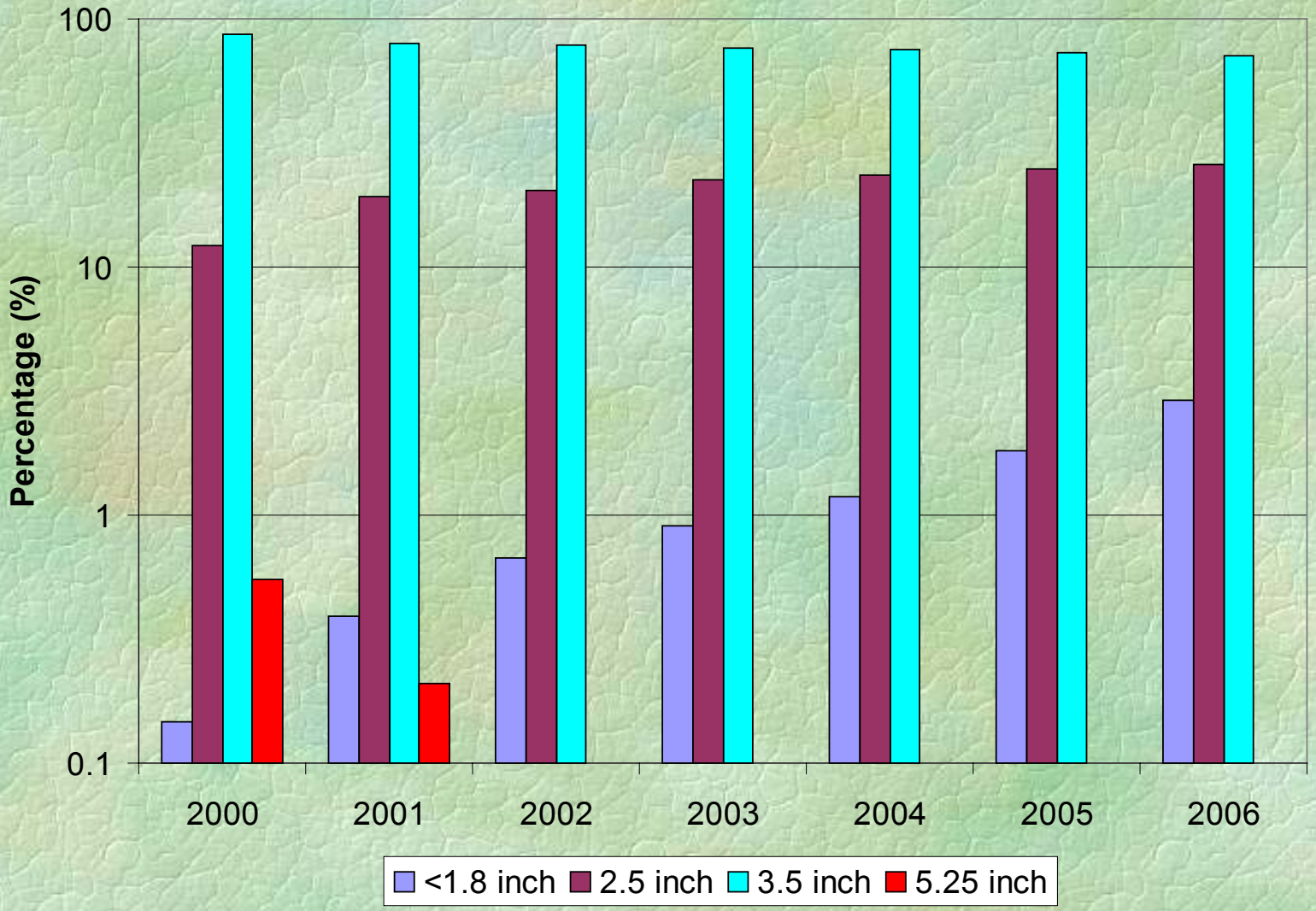
(Same Areal Density, 4 Disks)



Volumetric Density Comparison



Disk Drive Form Factor Changes

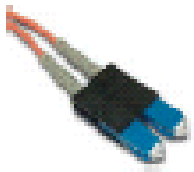


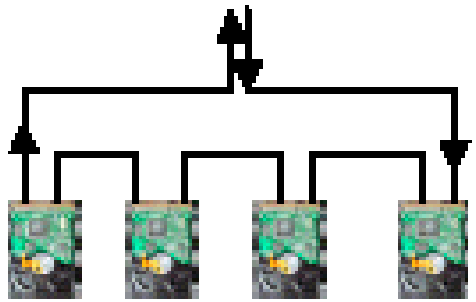
Today's Hard Disk Drive Interfaces

Typical Configuration

Key Features

Fibre Channel

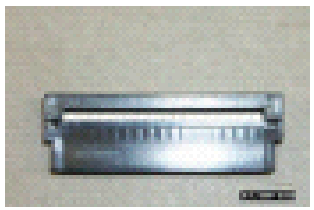


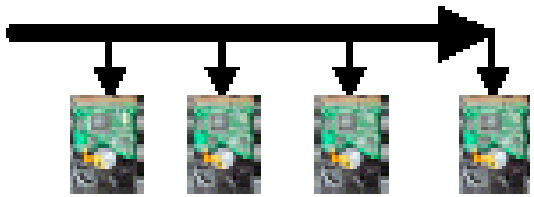


- Serial (low pin count)
- Point to point
- Looped
- Switched
- SCSI commands

Loop

SCSI

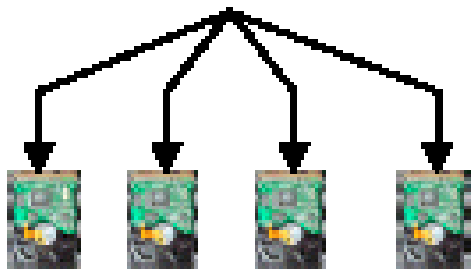




- Bus architecture
- SCSI Cmd. Protocol
- Reliability
- Expansion

Bus

Parallel ATA



- Low cost (thin protocol)
- Point to point interface

Point-to-point

Enterprise Storage Challenges

- **Key HDD requirements**
  - Reliability
  - Performance
  - Expansion
  - Hot Swapability
  - Back Plane drive capability (interface)
  - Cost
- **Parallel SCSI challenges**
  - Bus architecture is a bottle neck
  - Date Rate beyond 320 MB/s a huge challenge
  - 5V tolerance in < .15 micron silicon

Will SCSI really get replaced?

- **Fibre channel solves all the issues but…**
  - Both host side and drive side solutions come at a cost penalty
- **SATA addresses many of the Enterprise Storage requirements**

Drive Interface Migrations

Time

Parallel ATA → **Serial ATA**

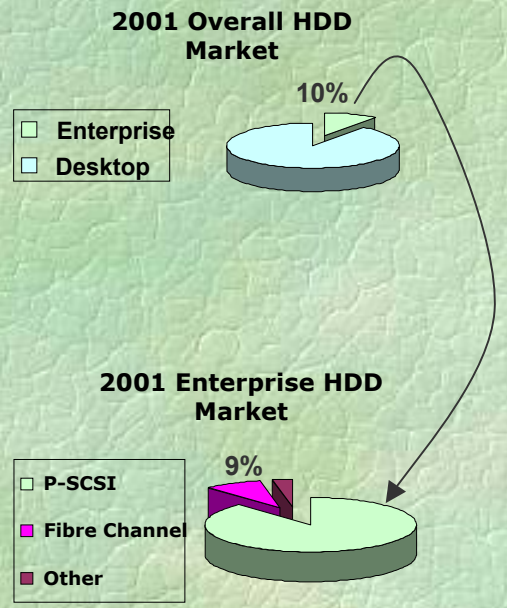
ATA is cost-optimized for non-mission critical applications

Parallel SCSI → **Serial Attached SCSI**

Serial Attached SCSI addresses the performance and reliability needs of enterprise environments

Fibre Channel → **Serial Attached SCSI & Fibre Channel**

Fibre Channel continues to pursue long-distance and connectivity solutions associated with SANs



Fibre Channel Speeds and Feeds

- 1 Gigabit per second (100 MB) since 1996
  - Physical layer adopted by Gigabit Ethernet
- 2 Gigabit per second (200 MB) since 1999
  - Gigabit Ethernet won't go there
- 4 Gigabit per second (400 MB) in 2003
  - Only a disk drive interface – not fabrics
- 10 Gigabit per second (1200 MB) in 2003
  - Physical Layer adopted from 10 Gig Ethernet

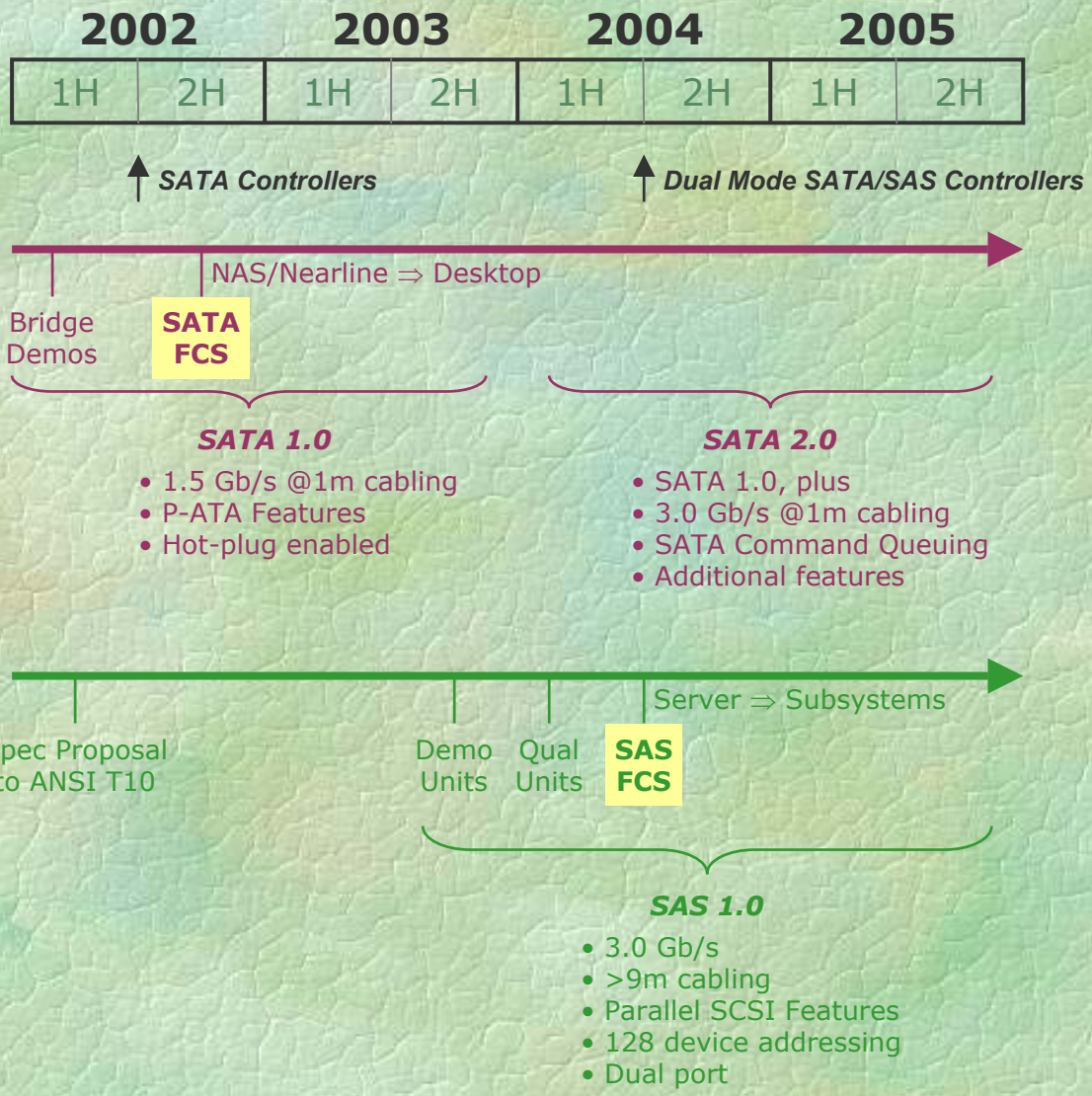
Interface Technology Comparison

	Serial ATA	Serial Attached SCSI
Performance	Half-duplex	Full-duplex with Link Aggregation
	1.5 Gb/sec (3.0 Gb/sec announced)	3.0 Gb/sec
Connectivity	Internal only	6m external cable
	One device	>128 devices
	No peer-to-peer	Peer-to-peer
Availability	Single-port HDDs	Dual-port HDDs
	Single-host	Multi-initiator
Driver Model	Software transparent with Parallel ATA	Software transparent with Parallel SCSI

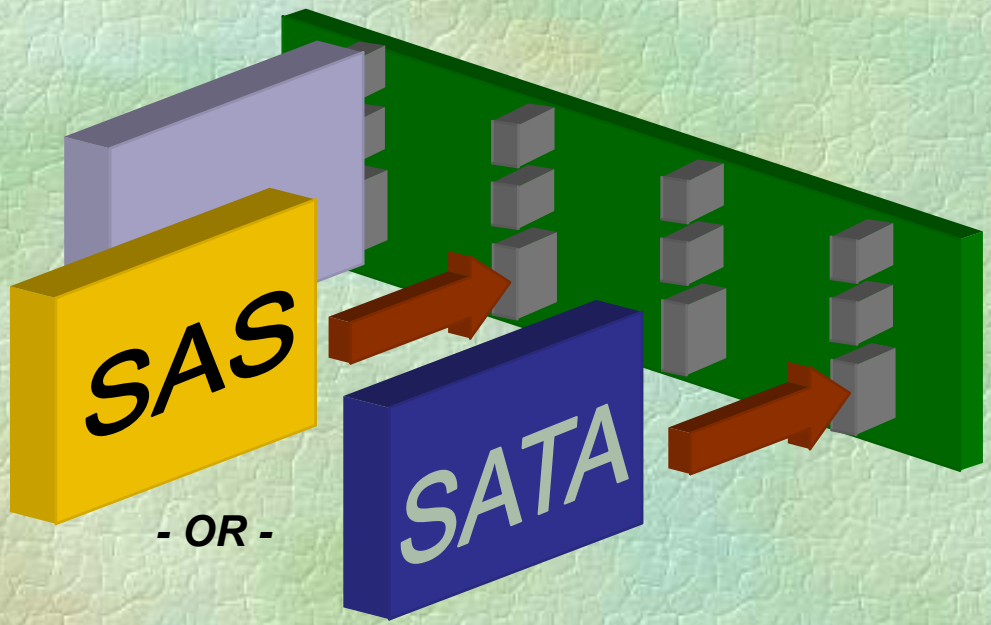
CE Interface Speed Comparison

	USB 2.0	1394	Serial ATA	Serial ATA Gen 2
Interface speed	*480 Mbps*	*400 Mbps*	*1500 Mbps*	*3000 Mbps*
Time to Copy 2GB File	40 sec	33 sec	11 sec	5 sec
Download 16 GB HD Movie	360 sec (6 min)	300 sec (5 min)	97 sec (1.6 min)	48 sec (0.8 min)
Back-up 80GB drive	1600 sec (27 min)	1333 sec (22 min)	427 sec (7.1 min)	213 sec (3.6 min)

General SATA & SAS Timelines



Enabling Choices For Customers



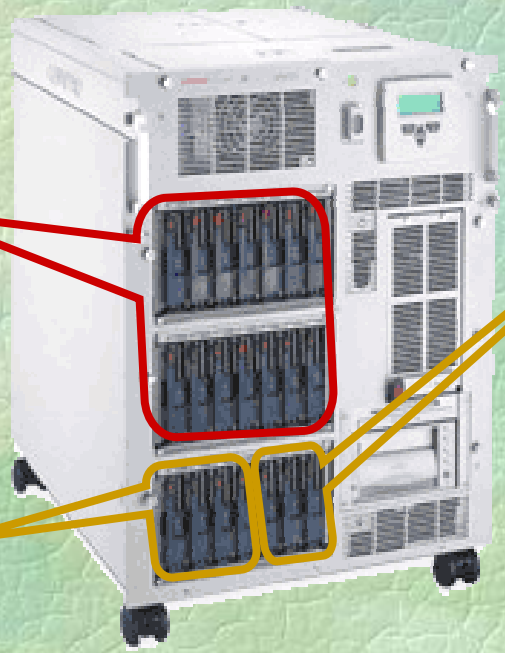
- A “properly designed” backplane can accommodate either SAS or SATA disk drives
  - SATA/High-Capacity disk drives can be used to enable “near-line” or tape augmentation applications
  - SAS/High-Performance disk drives can be used to enable “on-line” and performance-oriented applications
- Enables OEMS, VARs & Integrators the ability to re-use designs and more easily broaden their product offerings

Enabling Choices for Customers: *SATA-SAS Subsystem Example*

When drives can share a common controller & backplane, system designers & integrators are given more opportunities…

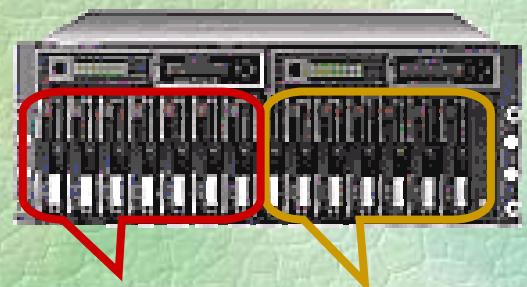
Dual port SAS drives for main stream server applications

SATA drives integrate disk to disk backup in the server to shorten backup and restore times



SATA drives with dual port, switched carriers for networked file storage

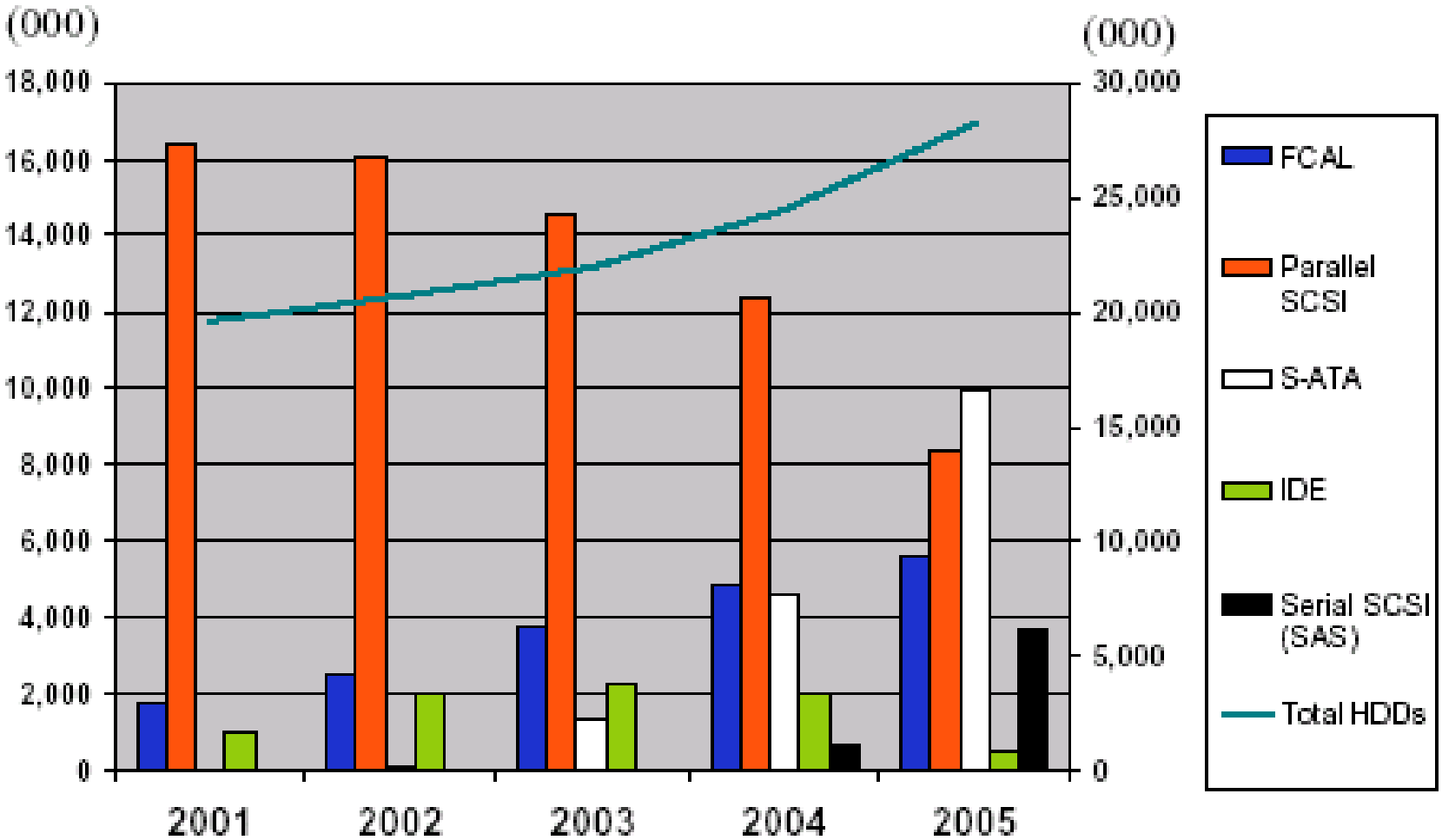
Add-on JBOD or RAID storage with mixed drive classes



SAS drives

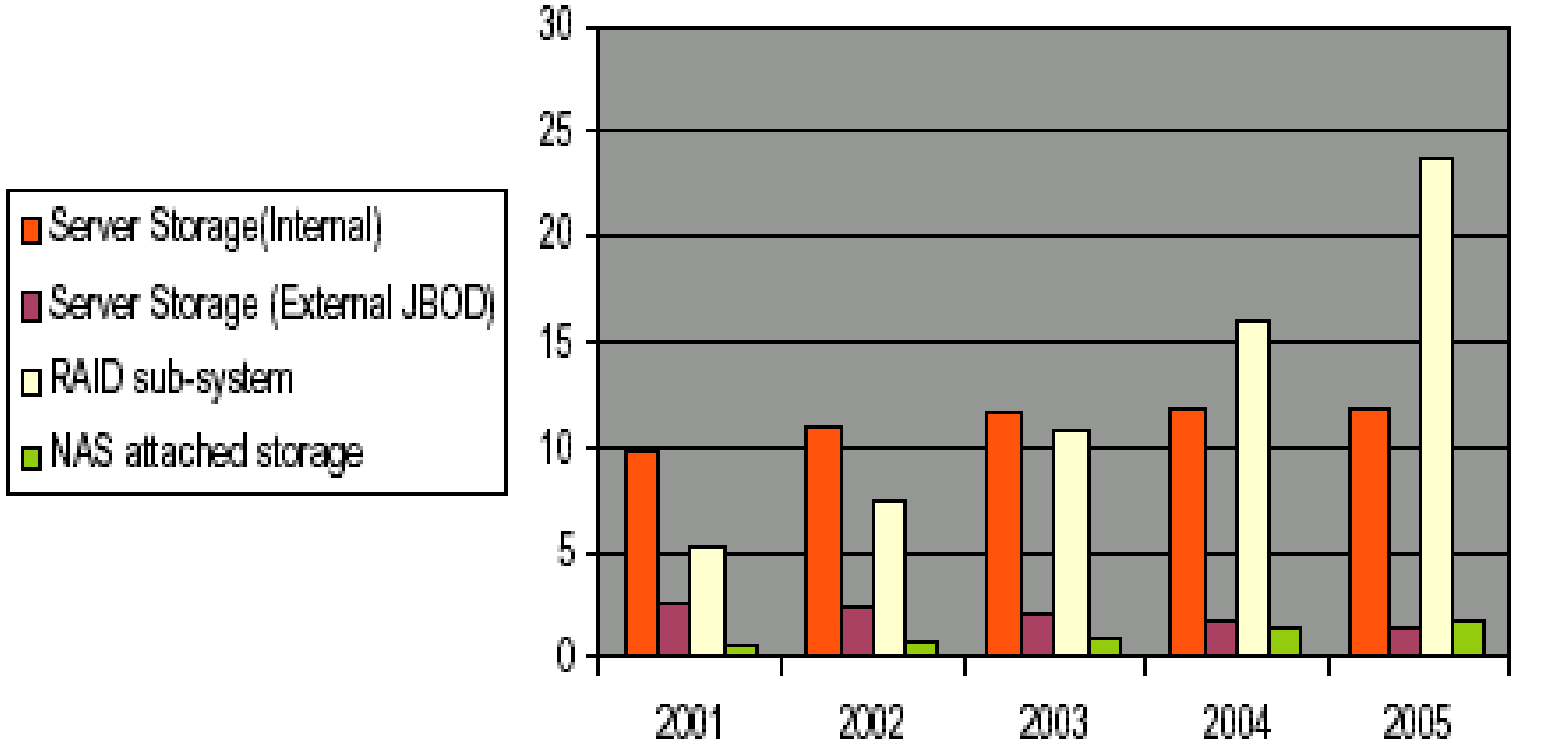
SATA drives

Enterprise Storage Interface Forecast



Source: IDC 3/02

HDD Unit Shipment Forecast by Segment



Source: Dataquest

Conclusions

- Data storage continues to grow. More things made digital.
- Greater need than ever to preserve our digital assets through backup and archive.
- Tremendous financial incentives tied to rapid recovery.
- Disk based backup will displace tape in many backup and restoration applications to create Enhanced Backup Storage.
- Three phases of Enhanced Backup Storage discussed, each leading to greater automation of backup and restore operations
- Changes in disk areal density and interfaces will lead to higher performance and less costly backup storage.
- Digital backup and archive remain a major component in data storage growth.