Almaden
Research Center

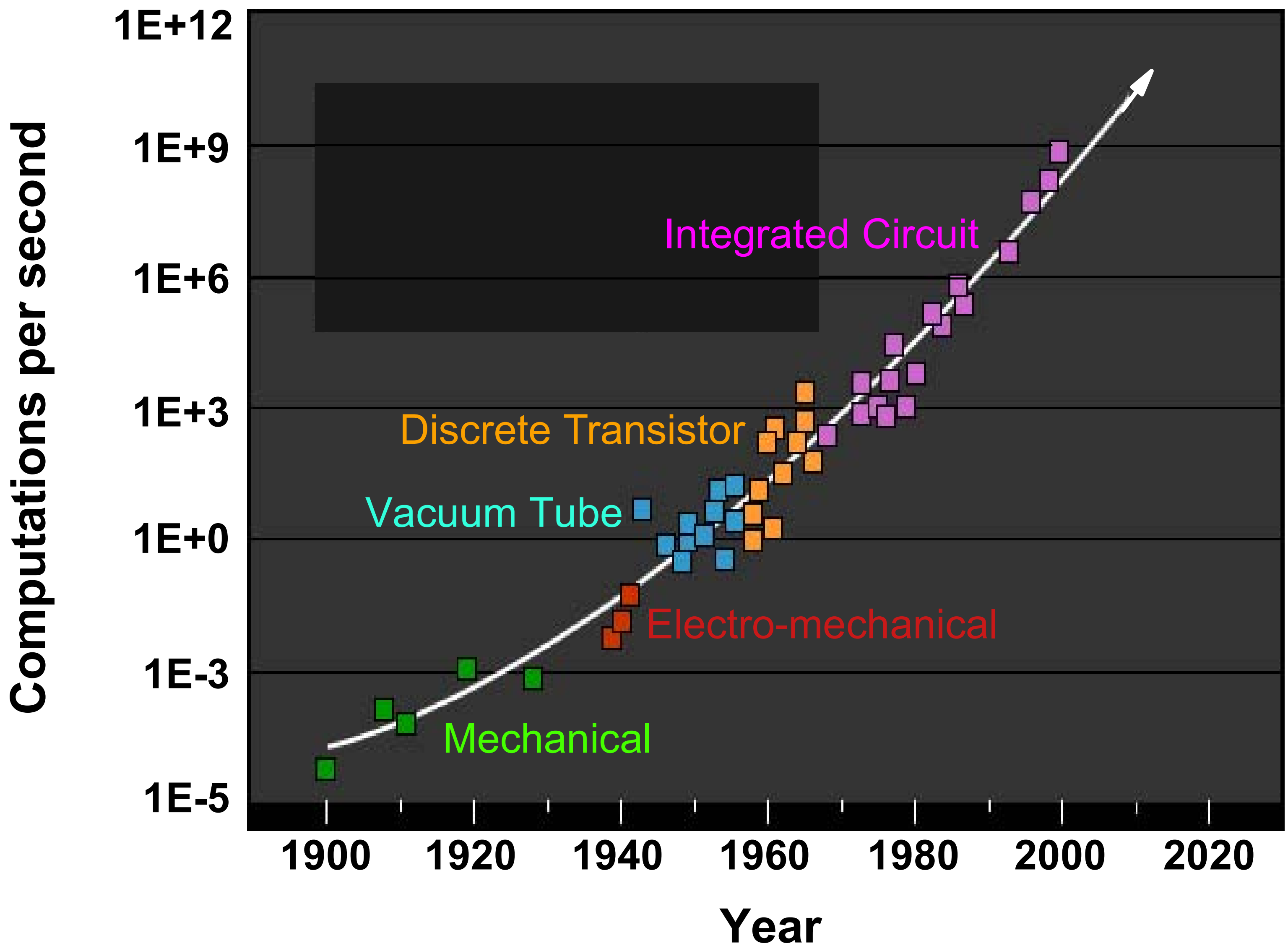
# Storage: From Atoms to People


Robert Morris

Director of Almaden Research Center
IBM Research

Almaden
Research Center

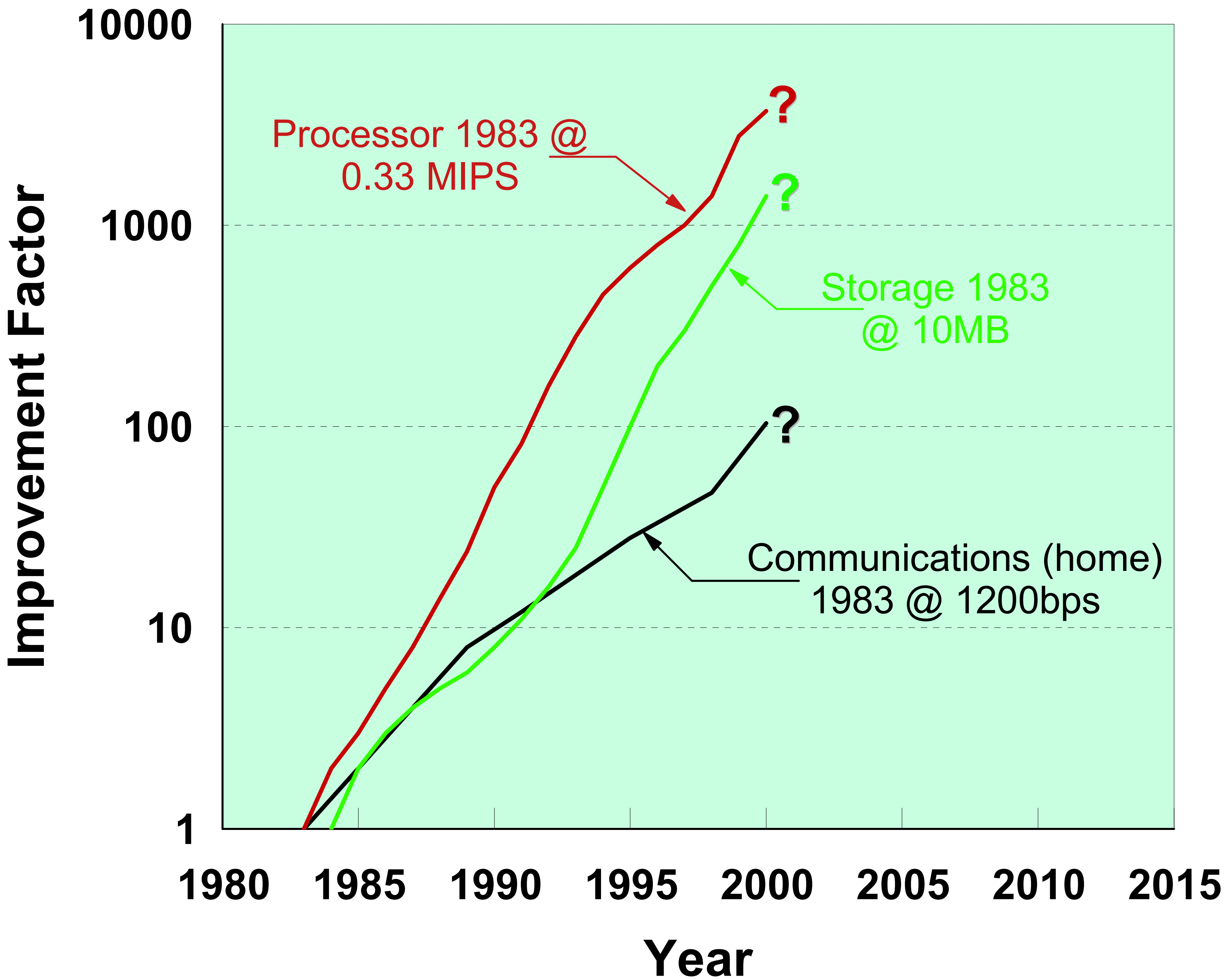
# $1000 Buys…




After "Mind Children: The Future of Robot and Human Intelligence," Hans Moravec, Harvard University Press, 1988 and R. Kurzweil, The Age of Spiritual Machines : When Computers Exceed Human Intelligence, Viking Penguin, 1999

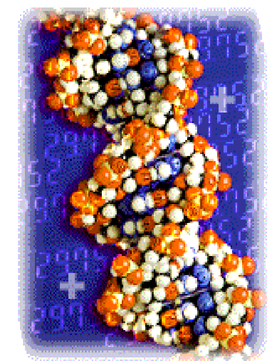
Almaden
Research Center

# Relative Trend of Technologies



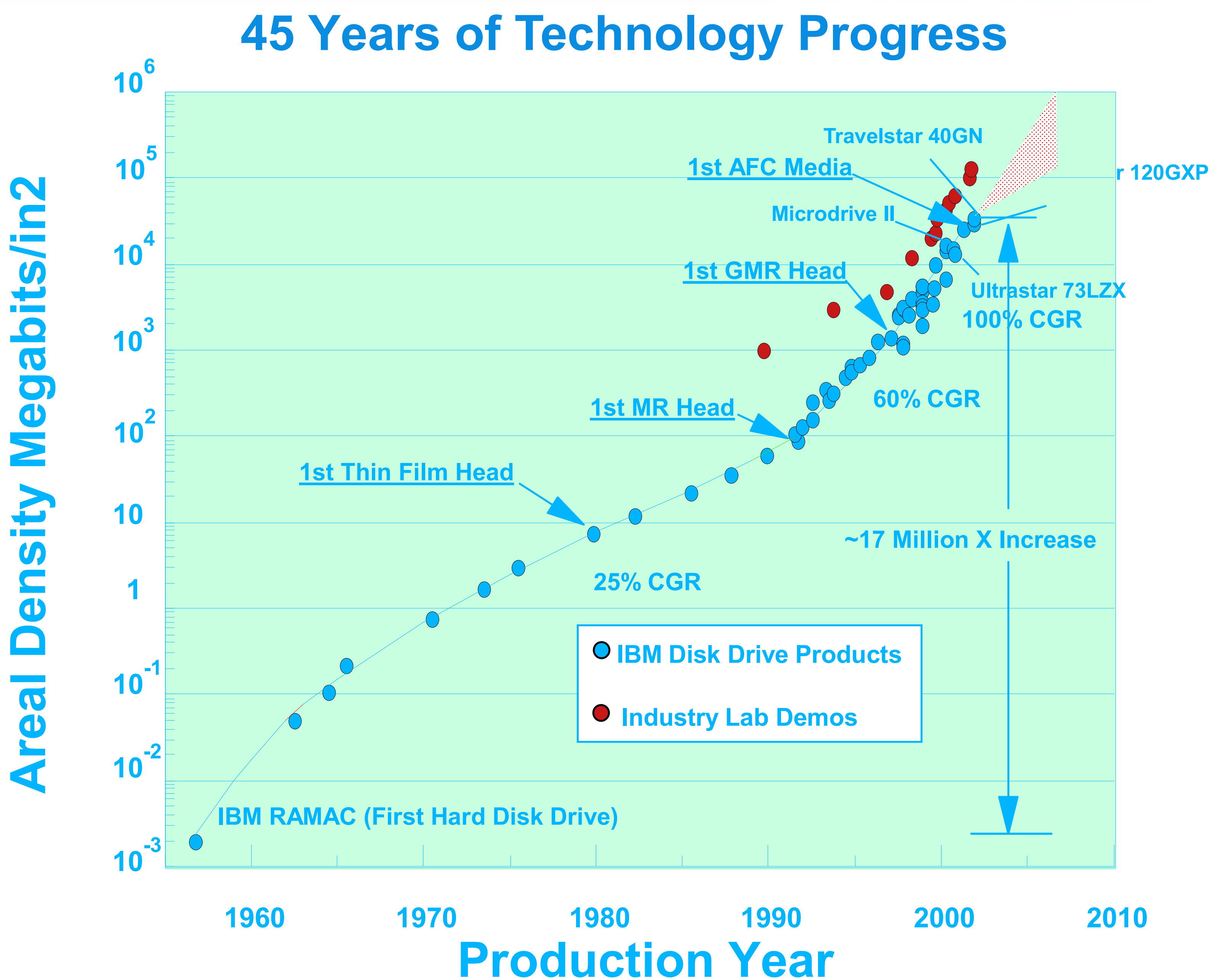

Source: IBM Research


# What Will We Carry?

| Carry Computer | Carry Storage | Carry Card | Carry Nothing |
|---|---|---|---|
|  |   |  |  |
| ThinkPad | Microdrive | SmartCard | Biometrics |

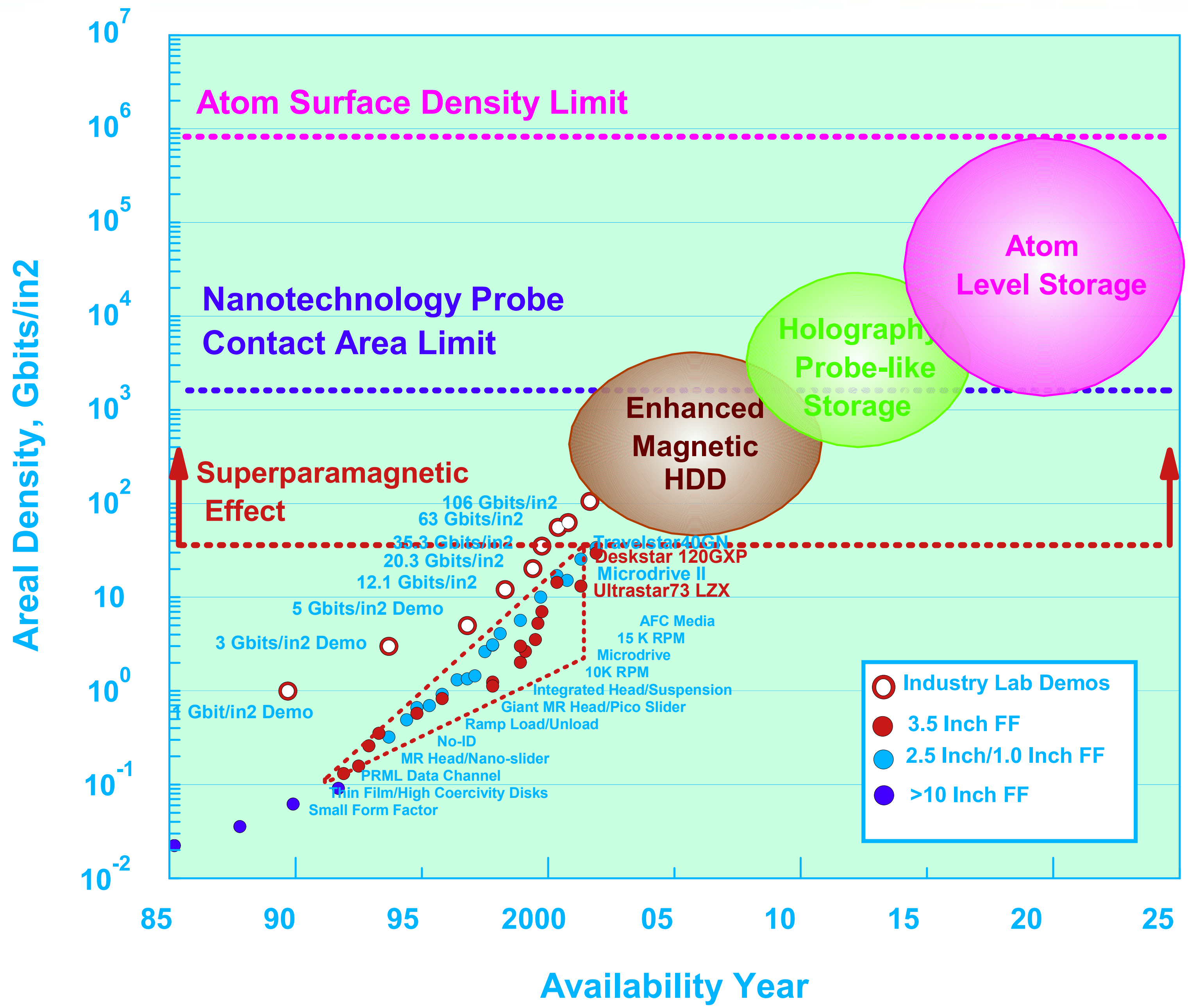
Almaden
Research Center

# HDD Areal Density Perspective



Almaden
Research Center

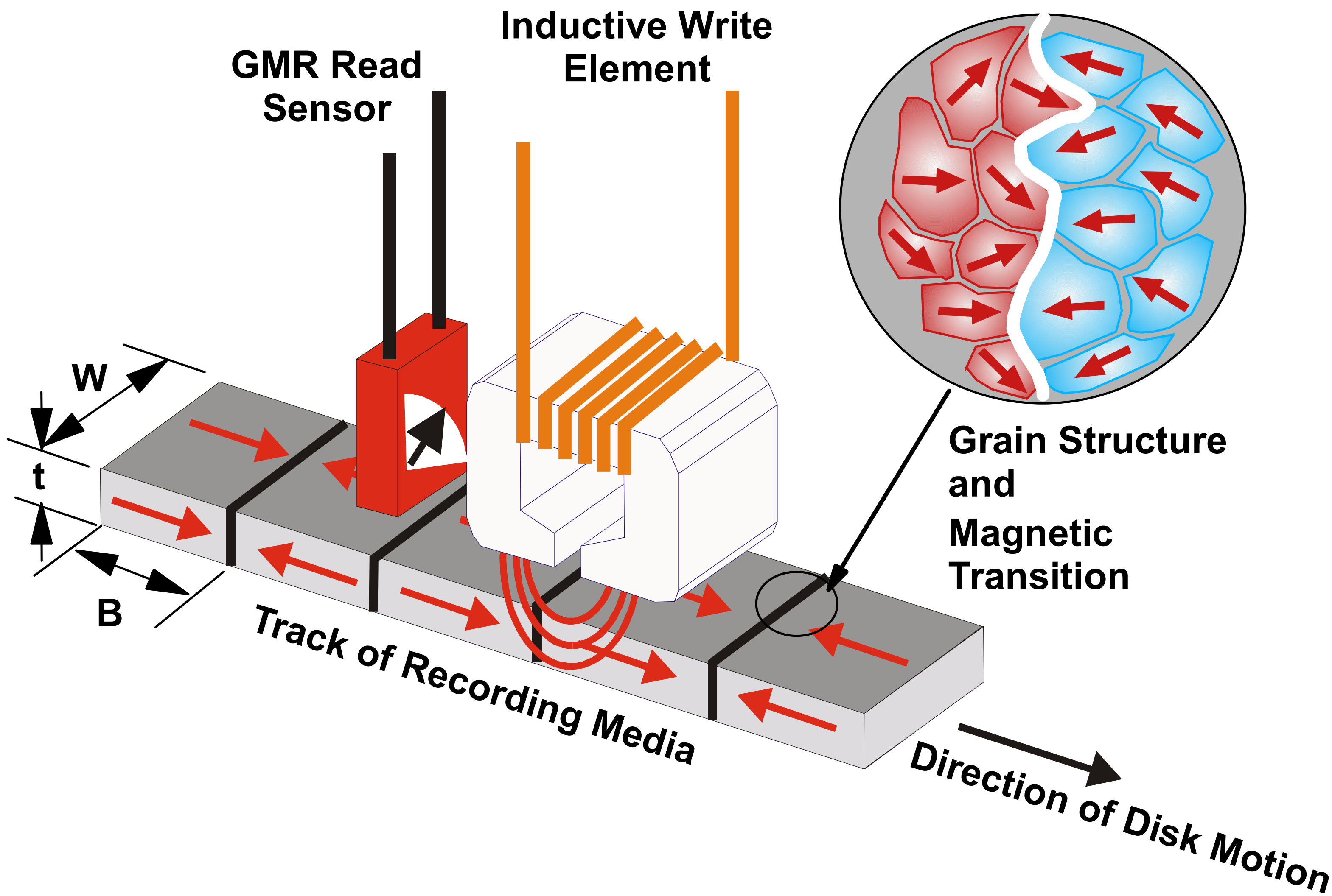
# Advanced Storage Roadmap




IBM

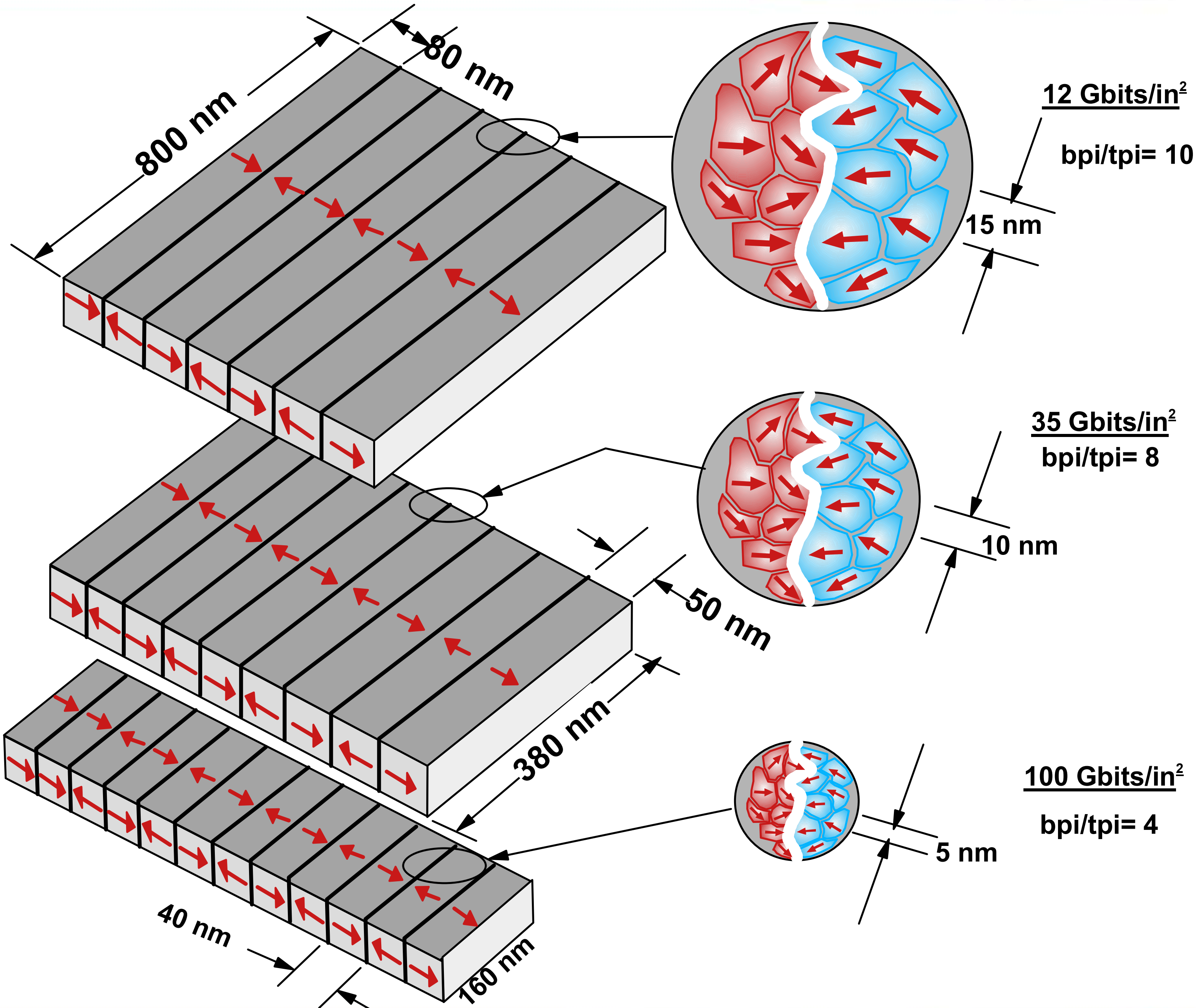
Almaden
Research Center

# Magnetic Recording Basics



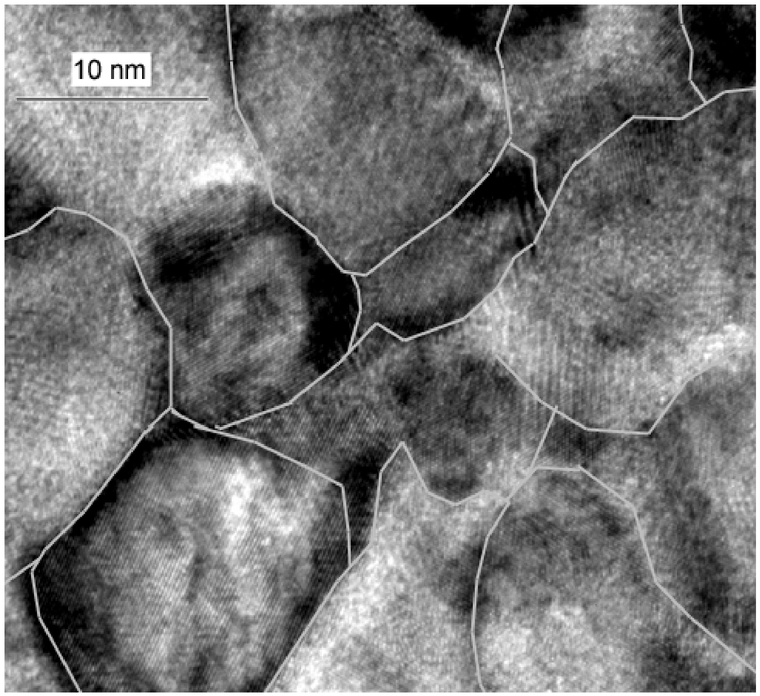
Almaden
Research Center

# Areal Density and Media Grain Size to Maintain ~1000 grains/bit

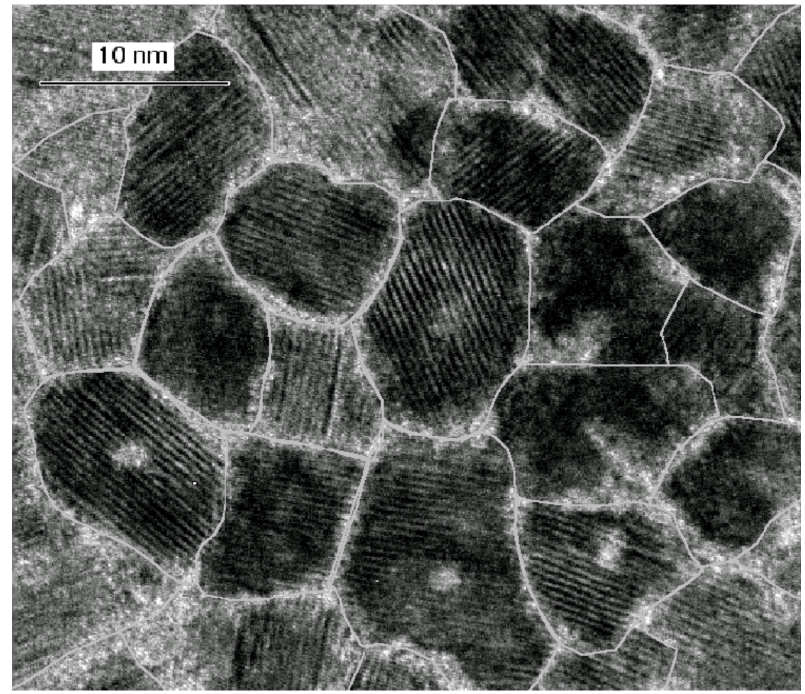



IBM

# Grain Structures in Magnetic Media



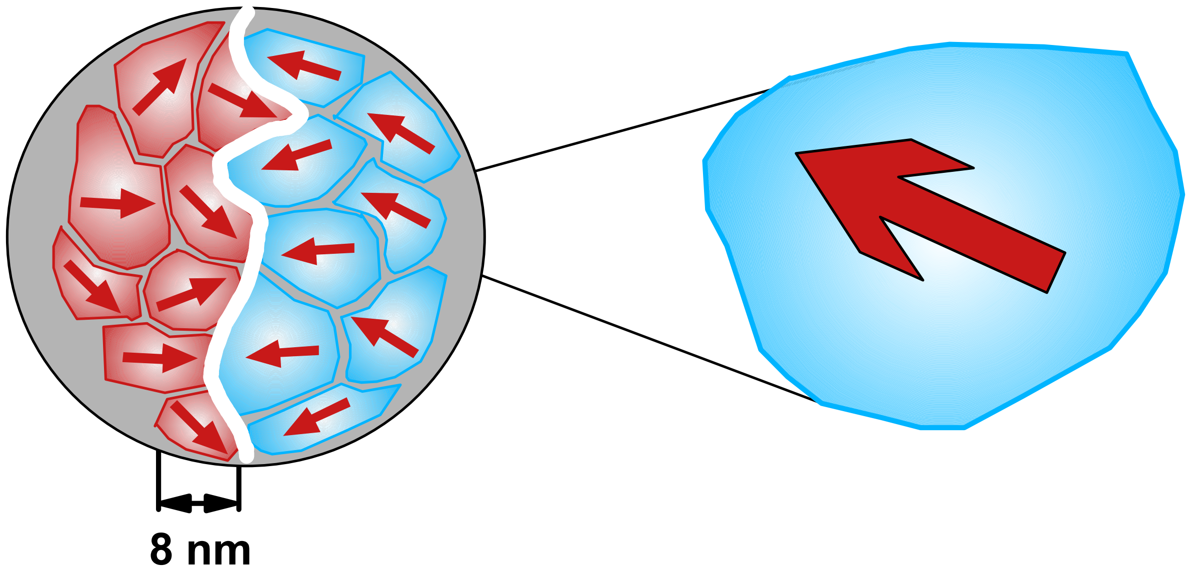

Areal density ~10 Gbits/in$^2$




Areal density ~ 25 Gbits/in$^2$

Magnification = 1 million

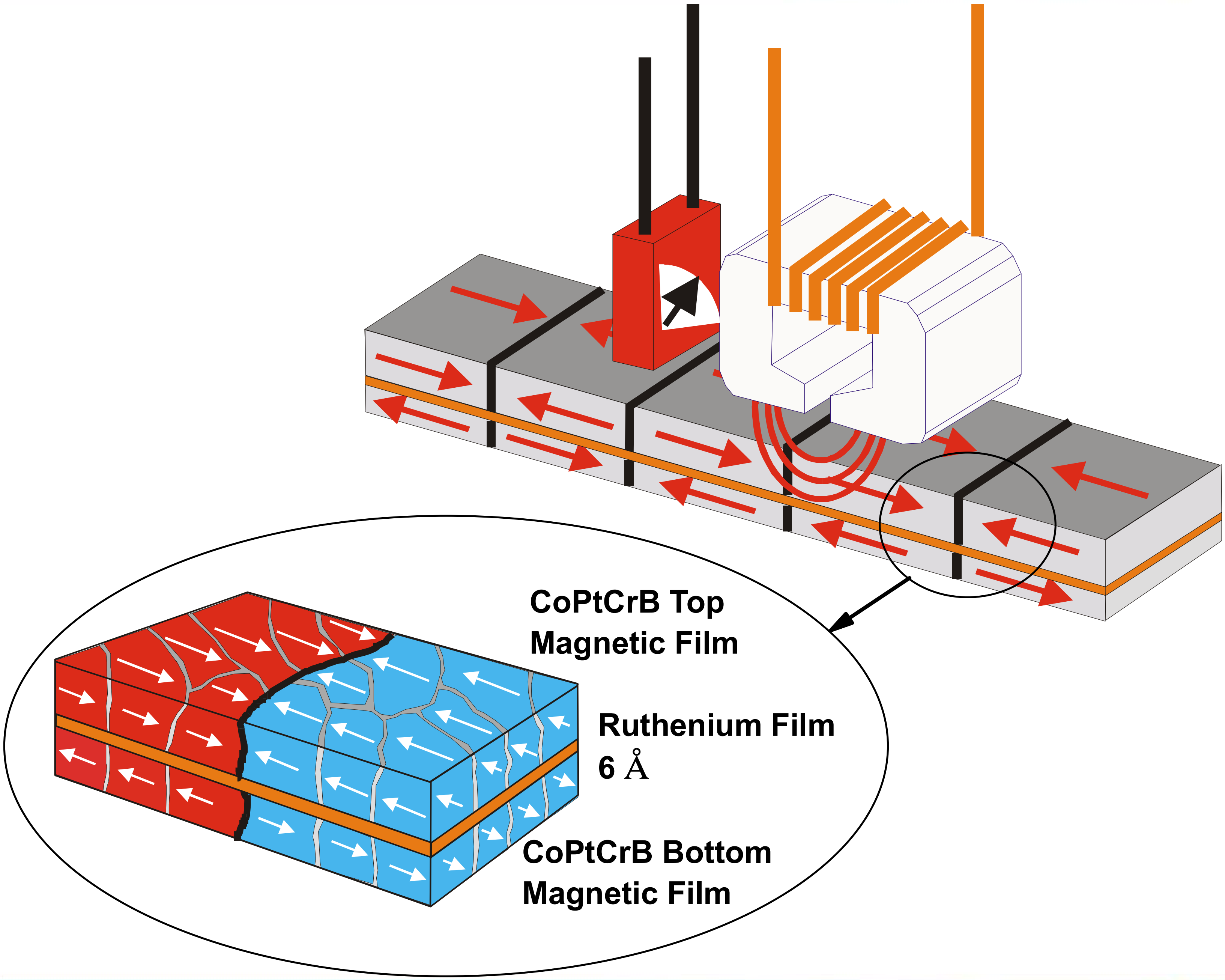
# Media Grain Size Scaling



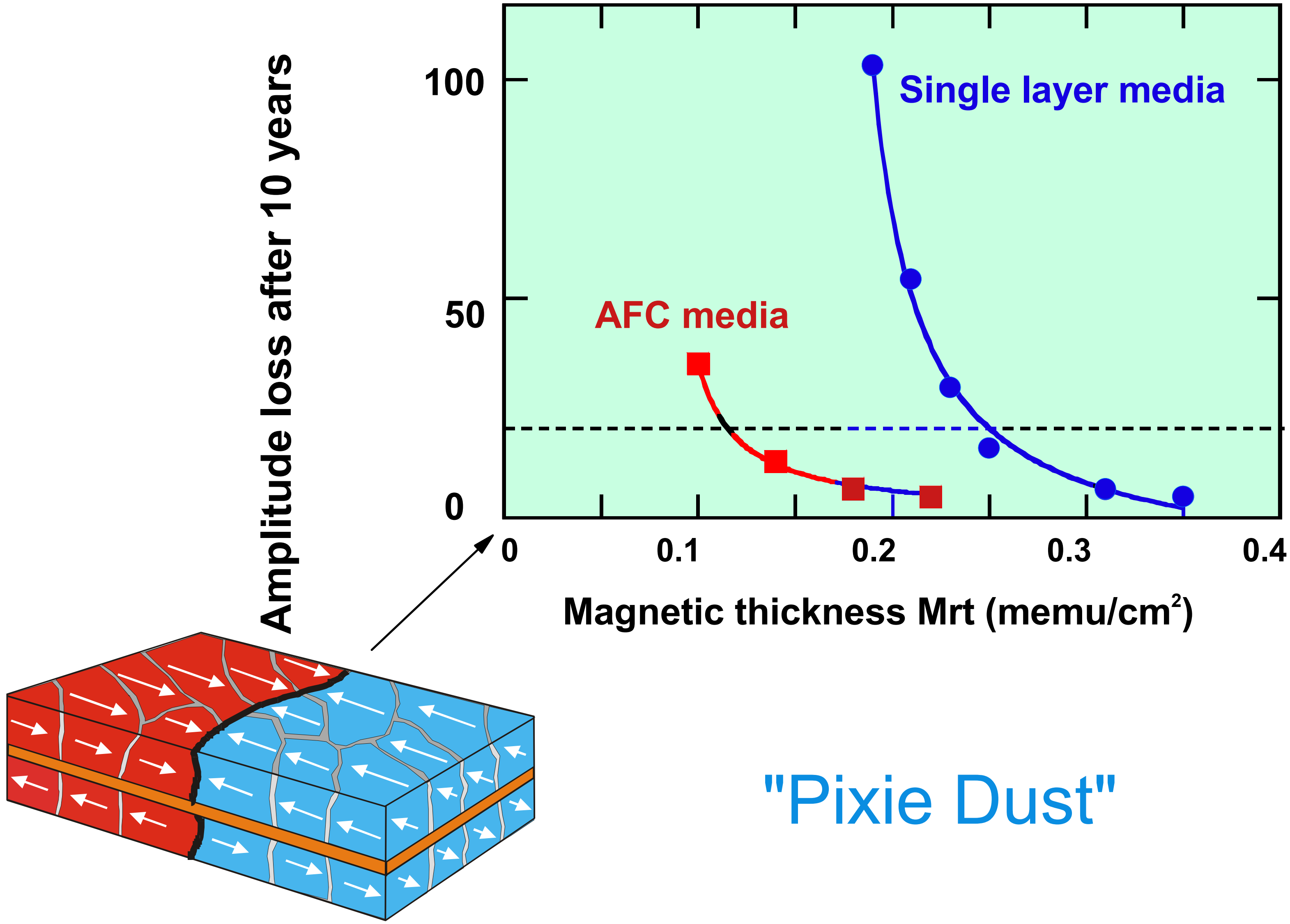

- Particle energy $E_{particle} \propto$ volume of grain
- Thermal stability requires that $E_{particle} > 55k_BT$ to store information for >10 years

**Superparamagnetic effect**

# Antiferromagnetically Coupled (AFC) Media Structure



# AFC Media Stability



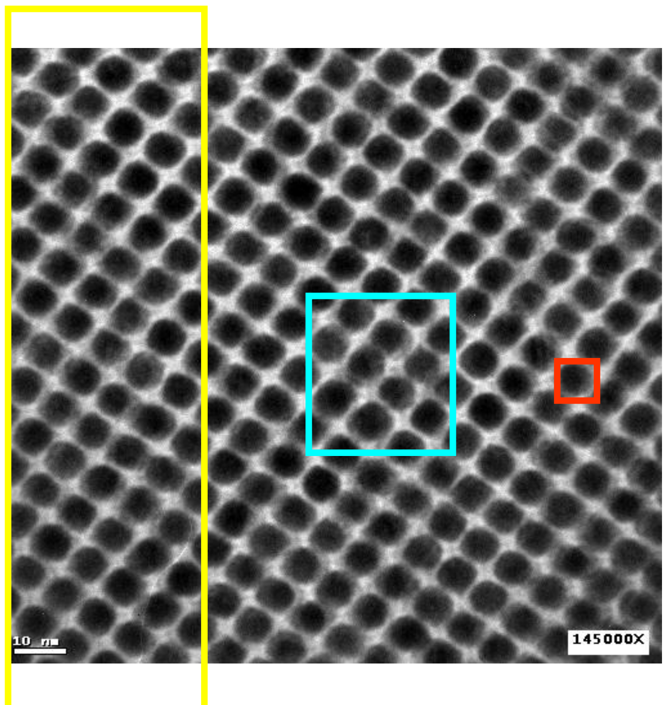

"Pixie Dust"

E. Fullerton, D. Margulies, M. Schabes, M. Carey, B. Gurney, A. Moser, M. Best, G. Zeltzer, K. Rubin, H. Rosen, M. Doerner,"Antiferromagnetically Coupled Magnetic Media Layers For Thermally Stable High Density Recording, Appl. Phys. Lett., 77, 3806 (2000).

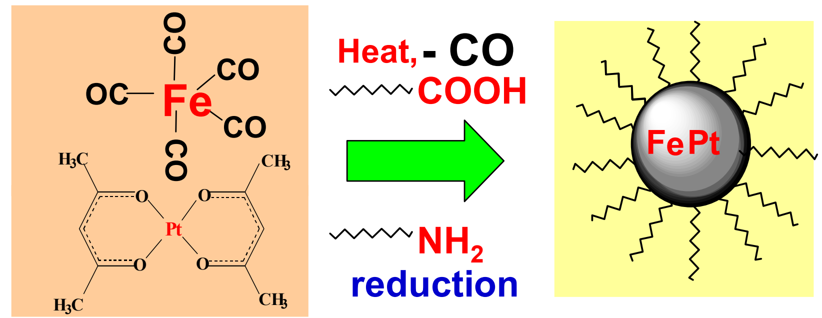
Almaden
Research Center

# Self Assembly for the Future

## Ultra High Density Magnetic Recording




| | | |
|---|---|---|
| 100 Gbit/in$^2$ bit cell ~130 particles 4:1 | 1 Tbit/in$^2$ bit cell ~13 particles 1:1 | 13 Tbit/in$^2$ bit cell ~1 particle 1:1 |



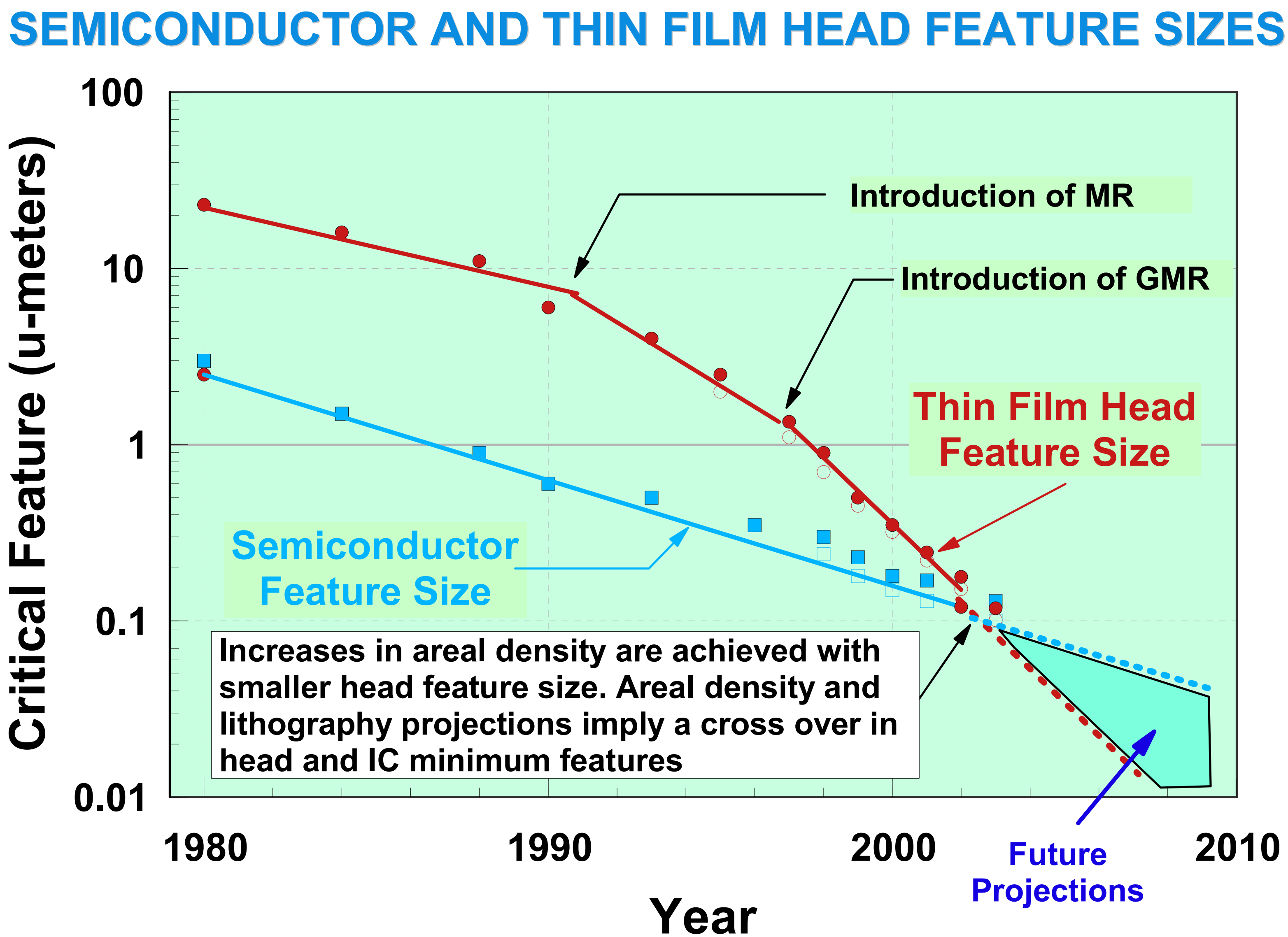

**Single magnetic domain per bit**
- Perpendicular media

**Challenge**
- Regular array over large area
- Adequate magnetic field to write particles

S. Sun, C. Murray, D. Weller, L. Folks, and A. Moser"Monodisperse FePt Nanoparticles and Ferromagnetic FePt Nanocrystal Superlattices", Science Vol. 287, 17 March 2000

Almaden Research Center

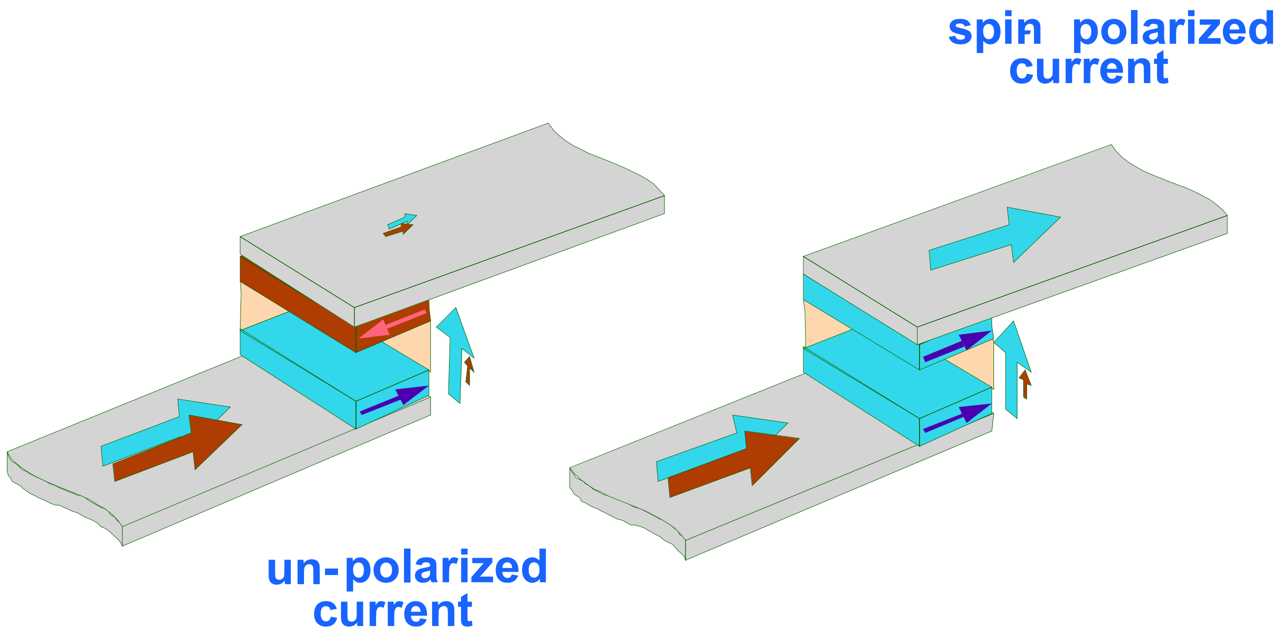
# Lithography Challenges

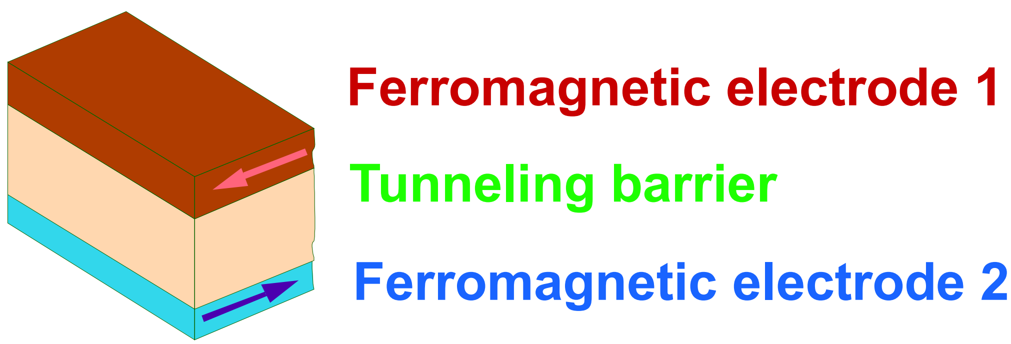



R. Fontana, J. Katine, M. Rooks, R. Viswanathan, J. Lille, S. MacDonald, E. Kratschmer, C. Tsang, S. Nyugen, N. Robertson, P. Kasiraj, To appear in IEEE Trans Mag.

IBM

# Magnetic Tunnel Junction



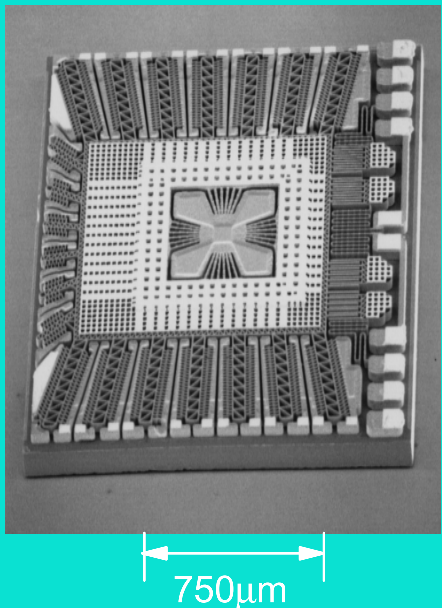




- First ferromagnetic electrode acts as spin filter
- Second FM layer acts as spin detector

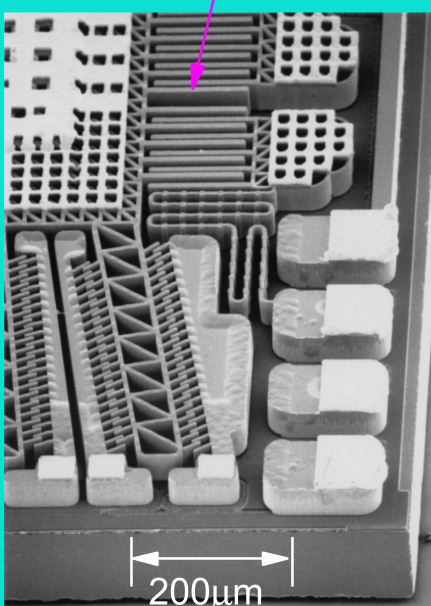
# Microactuator Technology for Track Following and Servoing

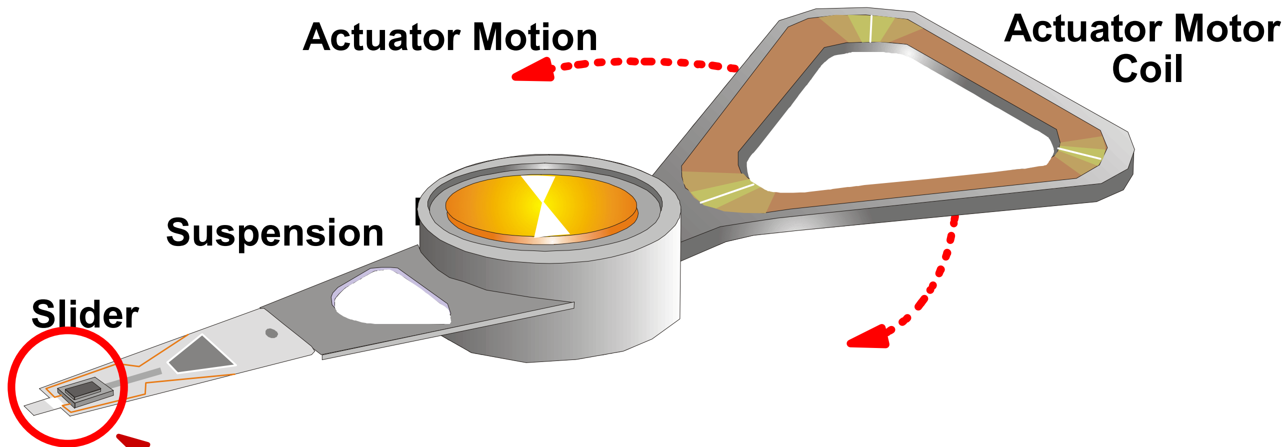
## Dual Stage Actuator using MEMS Technology

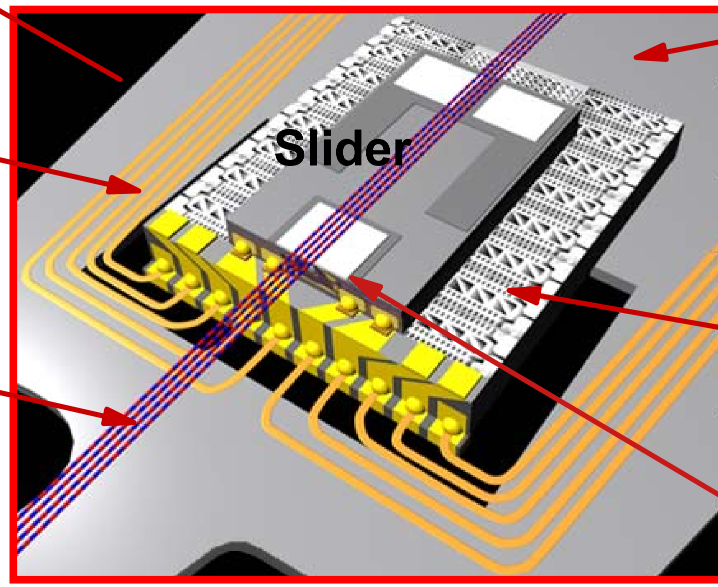
Overall View




Electrostatic Actuator



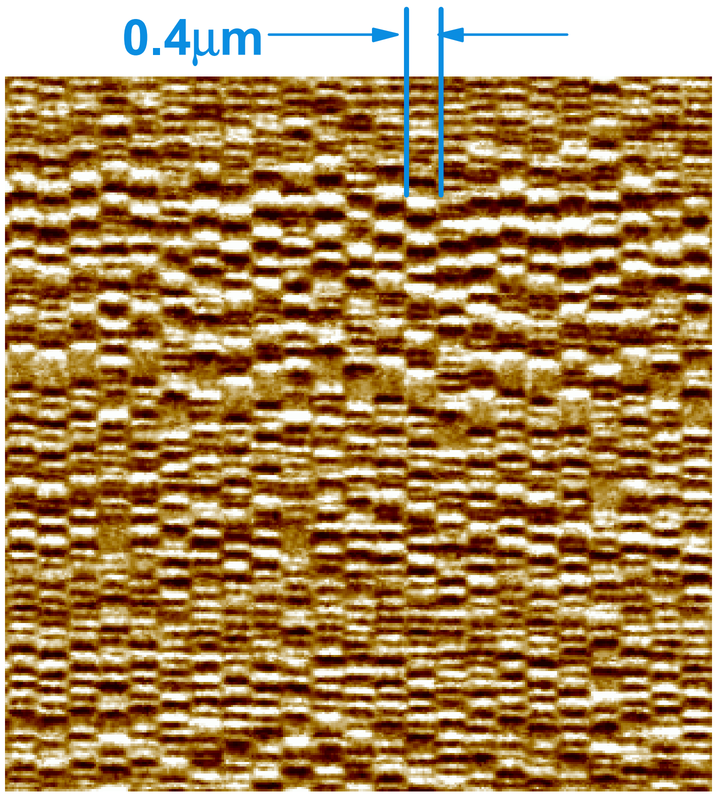




T. Hirano, M. White, X. Yang, T. Semba, V. Shum, S. Pattanaik, S. Arya, D. Kercher,and L. Fan, "3 kHz Servo Bandwidth Demonstration by HDD Tracking Microactuator, " Proc. 2001 ASME Int.l Mech. Eng. Cong. Nov. 2001.

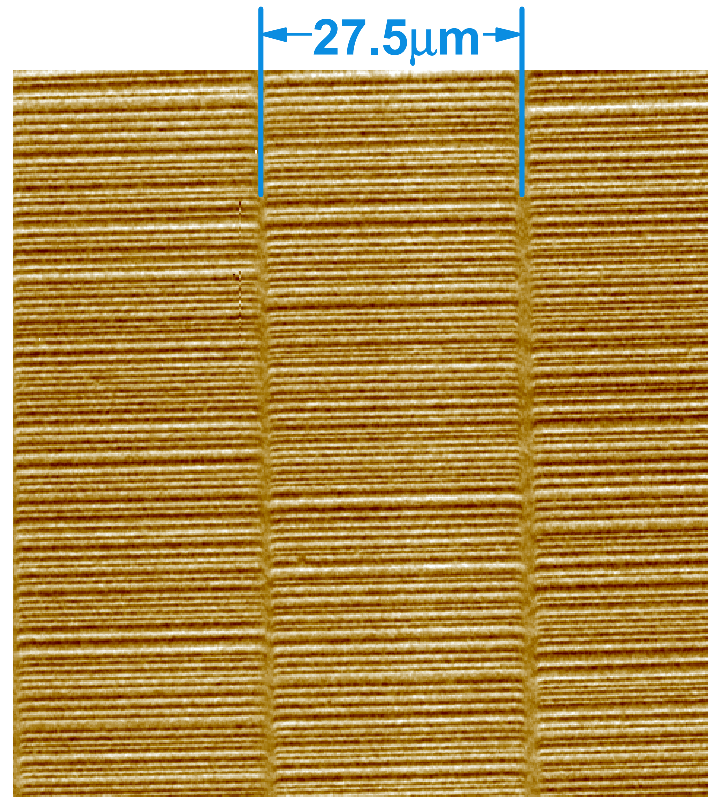

# Tape has Great Headroom for Growth

- Tape uses same basic magnetic recording technology as HDD
- Tape areal density is much lower than HDD

| | HDD | Tape | Ratio (HDD/Tape) |
|---|---|---|---|
| Bits per inch | 530,000 | 130,000 | 4 |
| Track per inch | 64,000 | 900 | 70 |
| Areal Density (Gb/in$^2$) | 34 | 0.1 | 280 |

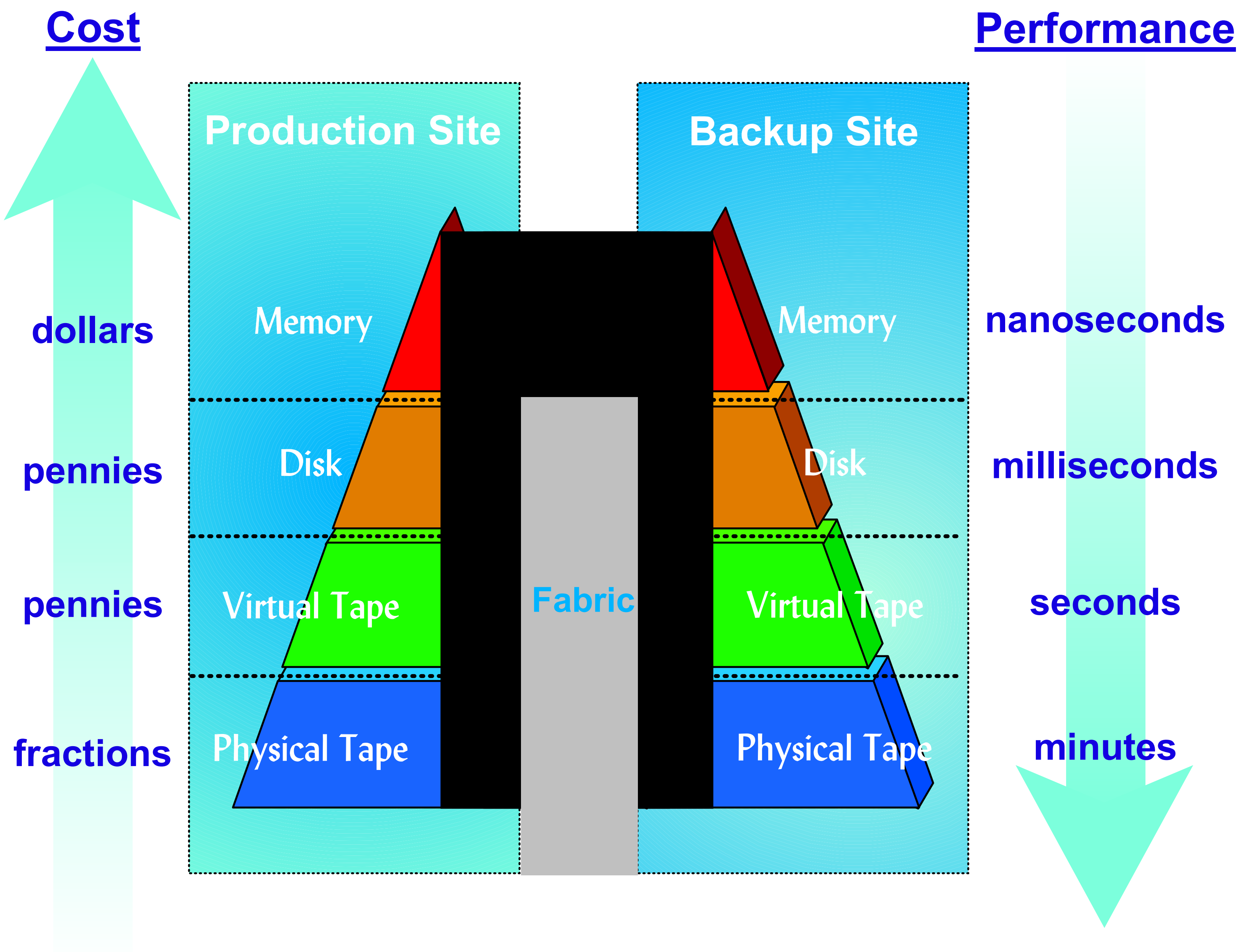



Mobile HDD (40GB/drive)




Tape: LTO (100GB/Cartridge)

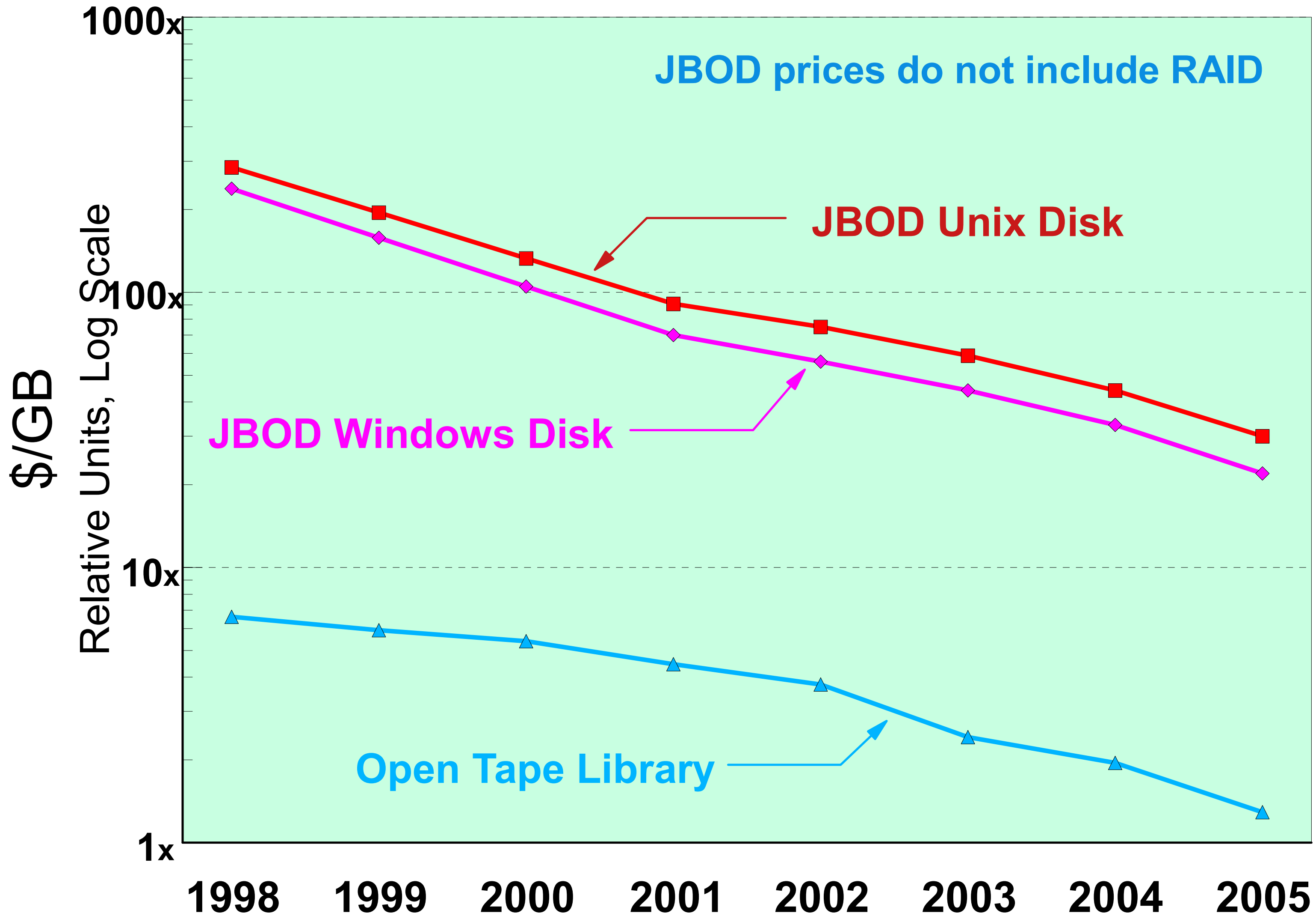
# Storage Infrastructure



Almaden Research Center

# Open Storage Subsystems

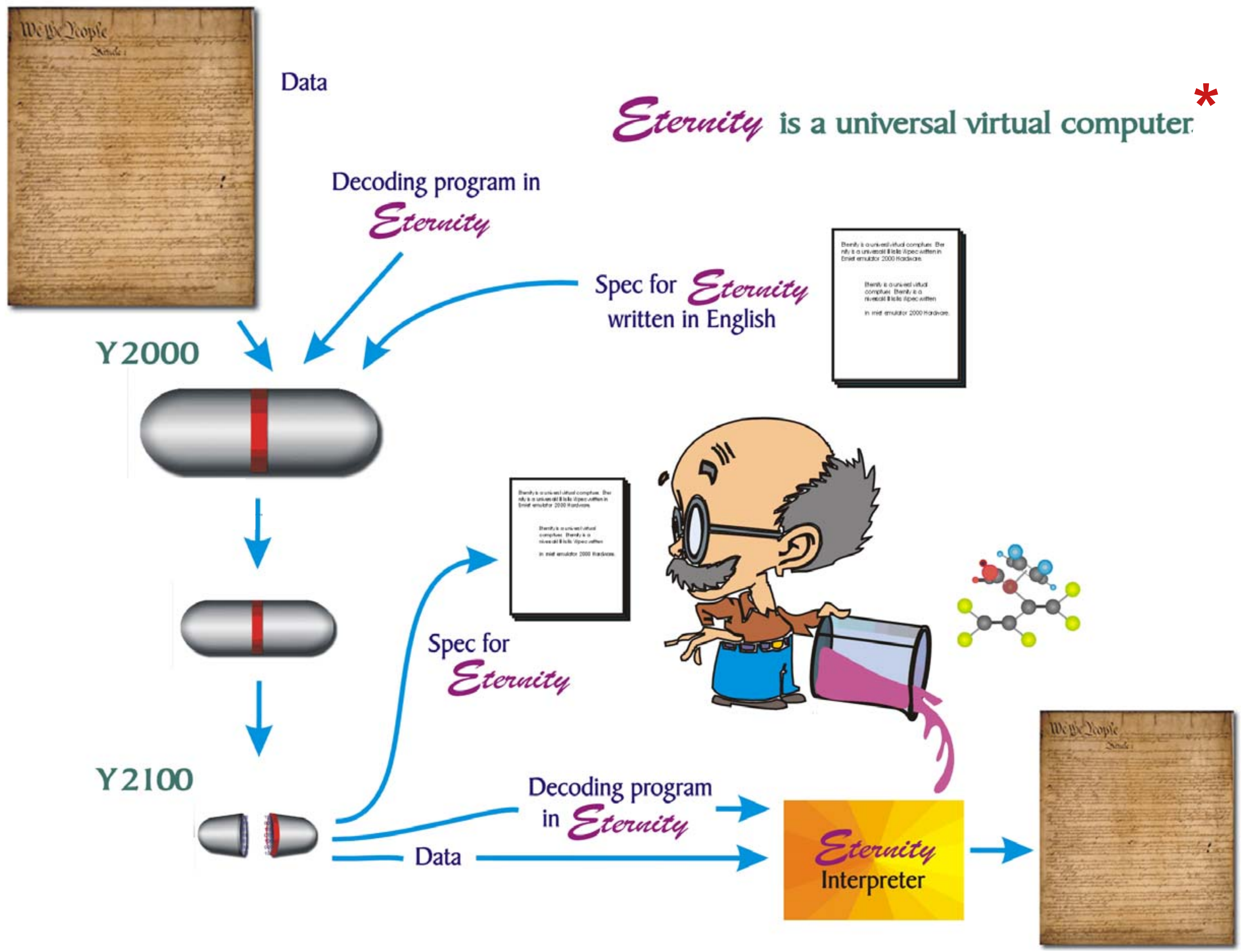
## Estimated Relative Price Trends




Source: Various IBM and Industry Studies

Almaden
Research Center

# An Approach to Data Preservation

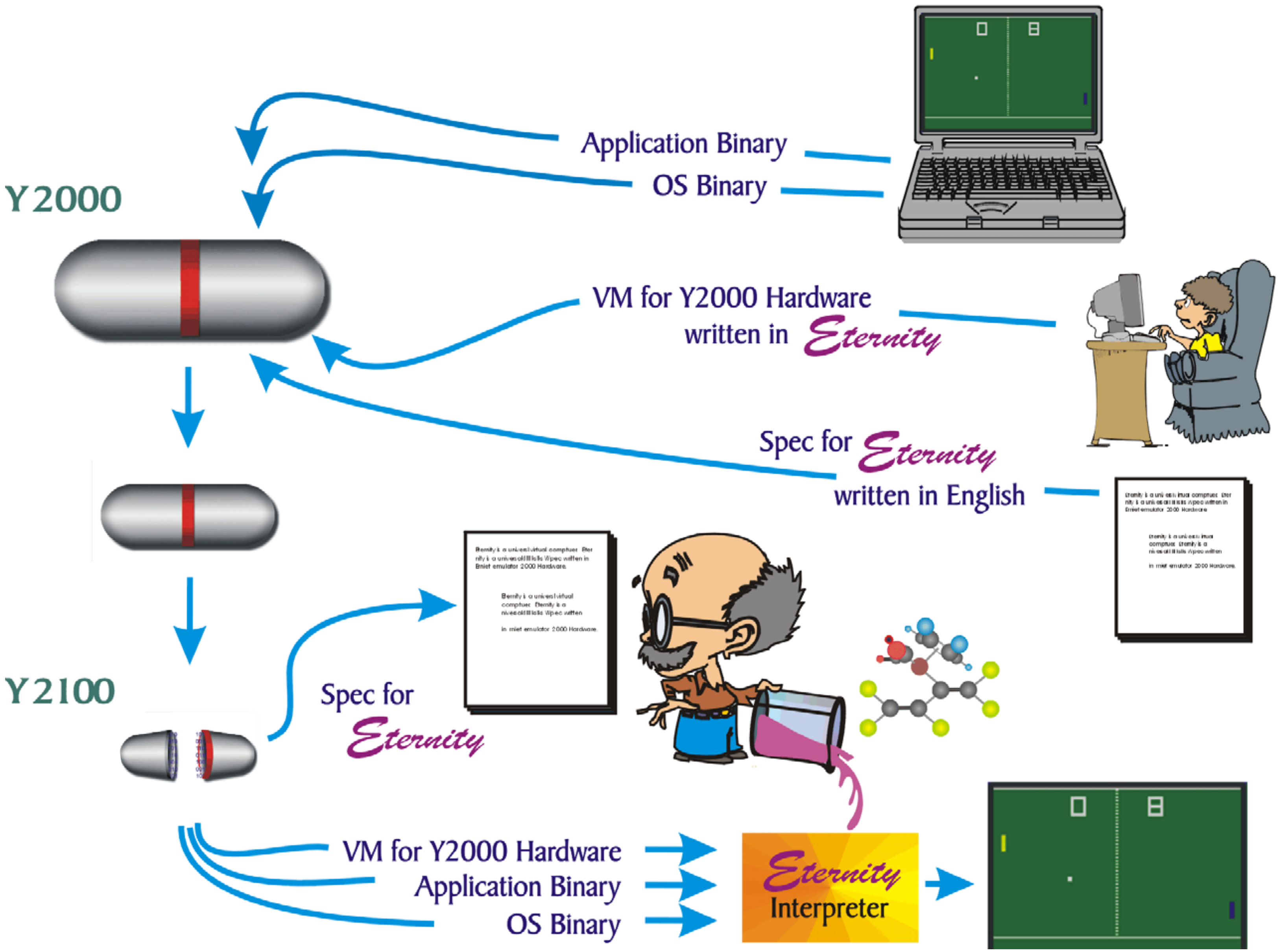



"Ensuring the Longevity of Digital Documents," by J. Rothenberg, Scientific American, 272 (1), January 1995.

*"Long Term Preservation of Digital Information," by R. A. Lorie, Presented at JCDL, May 2001.

Almaden
Research Center

# Preserving Progams as Well as Data



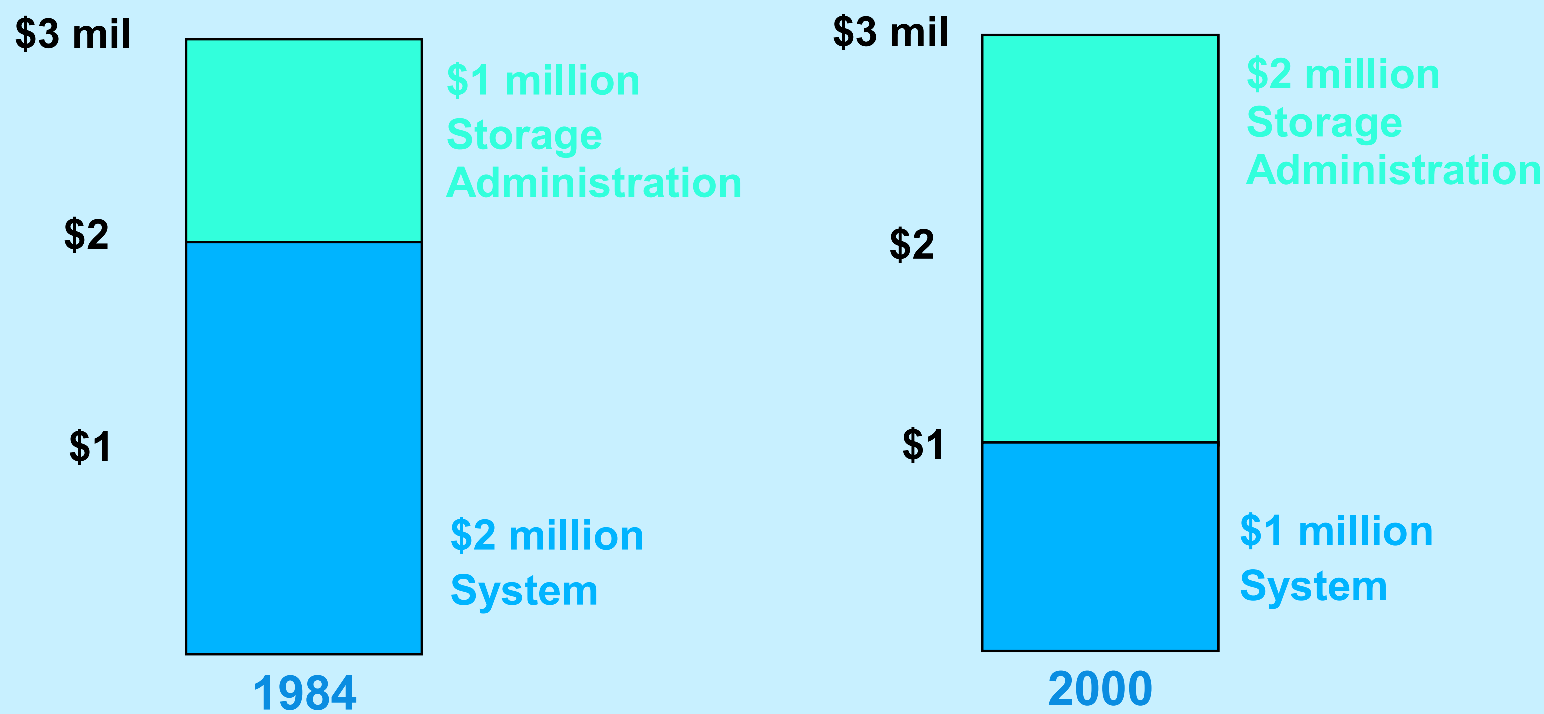

"Ensuring the Longevity of Digital Documents," by J. Rothenberg, Scientific American, 272 (1), January 1995.
"Long Term Preservation of Digital Information," by R. A. Lorie, Presented at JCDL, May 2001.

IBM

# The High Cost of I/T Management

## For example: the cost to manage storage is typically twice the cost of the actual storage system.

### Storage: What $3 million bought in 1984 and 2000.




(1) J. P. Gelb, "System-managed storage," IBM Systems Journal, Vol 28, No. 1, 1989 pp. 77-103.
(2) "Storage on Tap: Understanding the Business Value of Storage Service Providers", ITCentrix report, March 2001.
(3) "Server Storage and RAID Worldwide" (SRRD-WW-MS-9901), Gartner Group/Dataquest report, May 1999.

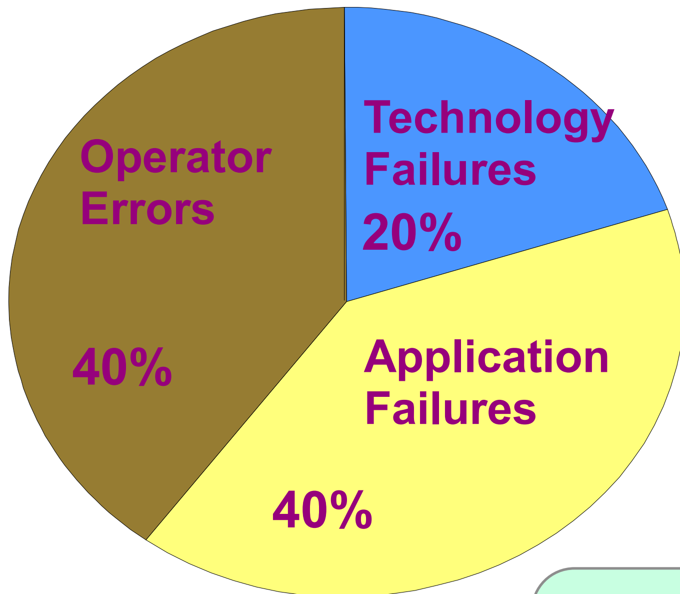
Almaden
Research Center

# Making the Front Page

**eBay**
Outage: 22 hours 12 June 1999
Operating System Failure
Cost: $3 million to $5 million revenue hit and 26% decline in stock price

**E*Trade**
3 February 1999 through 3 March 1999: Four outages of at least five hours
System Upgrades
Cost: ???
22 percent stock price hit on 5 February 1999

## Causes of Unplanned Application Downtime




**AT&T**
13 April 1998 outage: Six to 26 hours
Software Upgrade
Cost: $40 million in rebates
Forced to file SLAs with the FCC (frame relay)

**Dev. Bank of Singapore**
1 July 1999 to August 1999:
Processing Errors
Incorrect debiting of POS due to a system overload
Cost: Embarrassment/loss of integrity; interest charges

**America Online**
6 August 1996 outage: 24 hours
Maintenance/Human Error
Cost: $3 million in rebates
Investment: ???

**Charles Schwab & Co.**
24 February 1999 through 21 April 1999:
Four outages of at least four hours
Upgrades/Operator Errors
Cost: ???; Announced that it had made $70 million in new infrastructure investment.

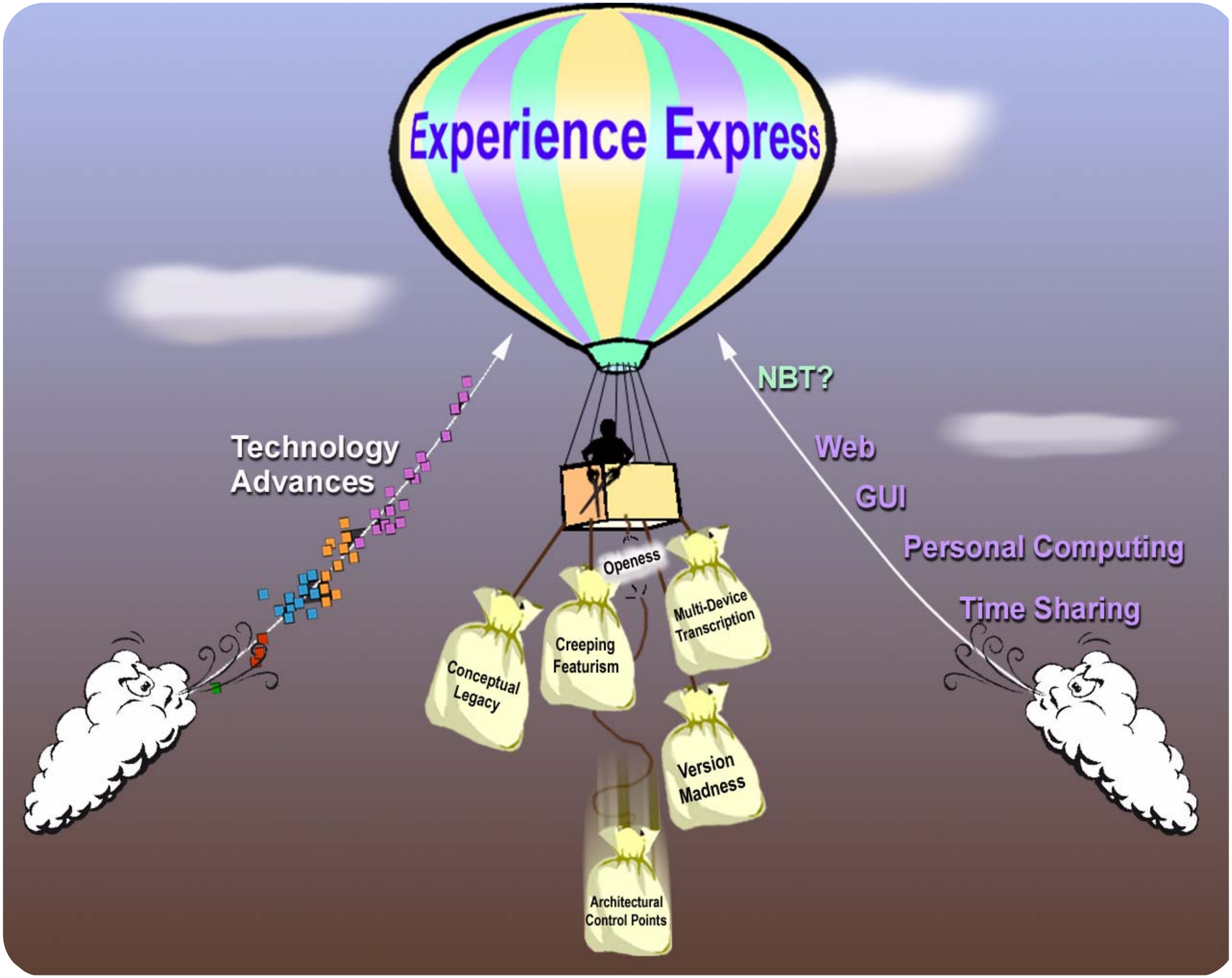
**NYSE**
June 8, 2001
>1700 stocks stopped trading for 90 minutes
Software Upgrade
Cost: ???

Various sources including Gartner Group

IBM®

Almaden Research Center

# (R)evolution of User Experience

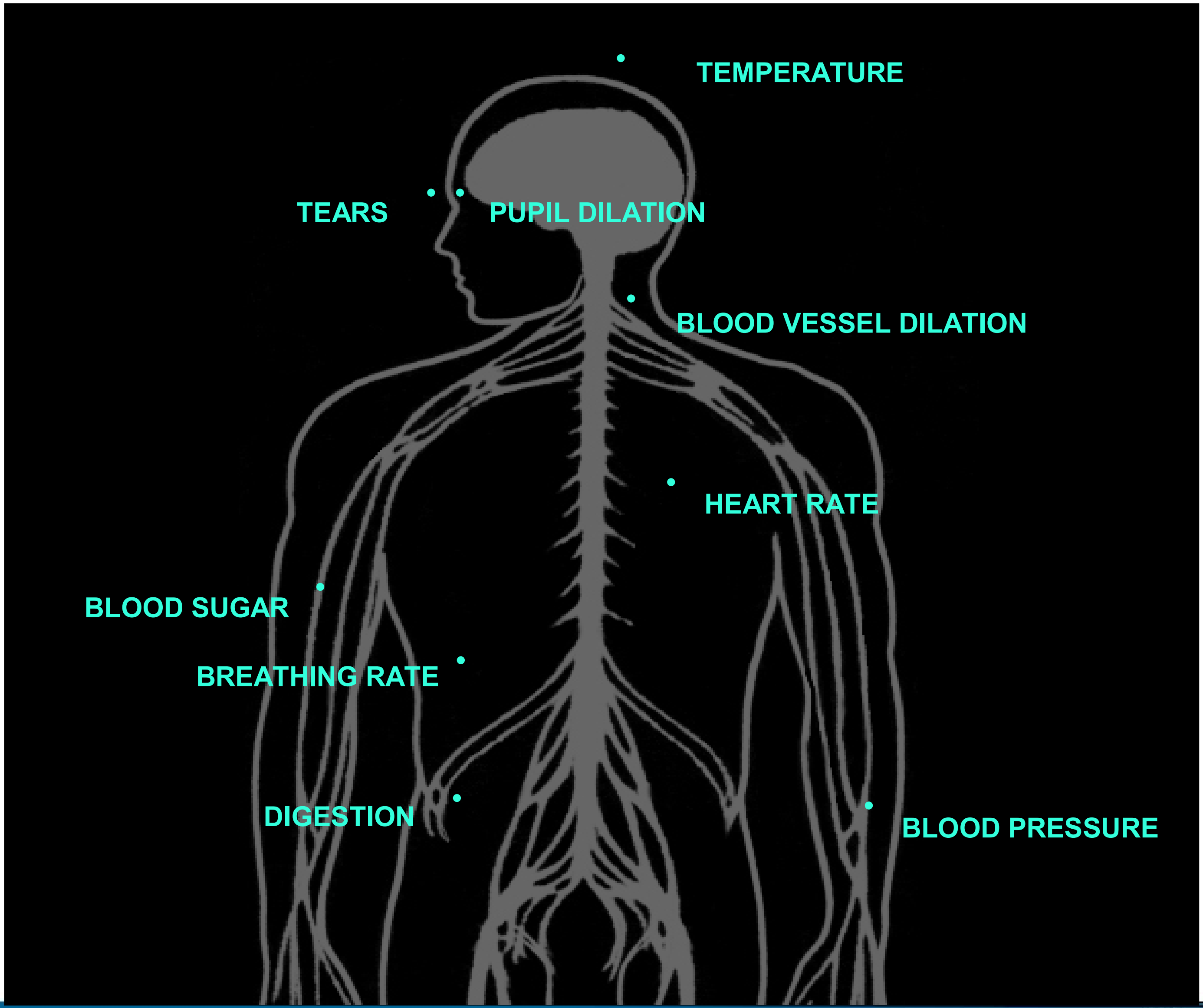



IBM

Almaden Research Center

# Autonomic Computing

http://www.research.ibm.com/autonomic



# Autonomic Computing

**Self-defining:** A system's understanding of its make-up, parameters and connections with other systems.

# Autonomic Computing

**Self-defining**

**Self-configuring and Self-optimizing:** The system's ability to adjust to its configuration and resource allocation to achieve predetermined goals.


# Autonomic Computing

**Self-defining**

**Self-configuring and Self-optimizing**

**Self-healing and Self-protecting:** **The system’s ability to anticipate and respond to attacks and failures by reallocating workflow or shifting specific functions to achieve stability.**


# Autonomic Computing

**Self-defining**

**Self-configuring and Self-optimizing**

**Self-healing and Self-protecting**

**Contextually Aware in a Heterogeneous Environment:**
**The system's ability to work seamlessly with other systems and adjust its actions based on context.**


# Autonomic Computing

**Self-defining**

**Self-configuring and Self-optimizing**

**Self-healing and Self-protecting**

**Contextually Aware in a Heterogeneous Environment:**

**Anticipatory: The system’s ability to anticipate workflow challenges and optimize the system for a user's immediate needs.**

Almaden
Research Center

# Autonomic Computing Evolution

| Levels of Autonomic Computing Sophistication | Well known examples | Current Research and Product Directions | Future Goal |
|---|---|---|---|
| Serving the World (people, business processes) | SMS | Storage Tank | Future Innovations |
| Heterogeneous Components Interacting | SNMP | Future Innovations | Future Innovations |
| Homogeneous Components Interacting | Adaptive network routing, Network congestion control<br>High availability clustering | Collective Intelligence Storage Bricks<br>Oceano | New packaging concepts for storage<br>Subscription computing |
| Components | ESS<br>RAID<br>DB Optimizer<br>Virus Management | Regatta self-healing, LPAR<br>SMART/LEO<br>Software Rejuvenation | More of the same and better |

IBM


# Storage Systems

## Key Trends:

- **Virtualization**
- **Self-Management**
- **Modularity**
- **Fail-in-place**
- **Policy Management**

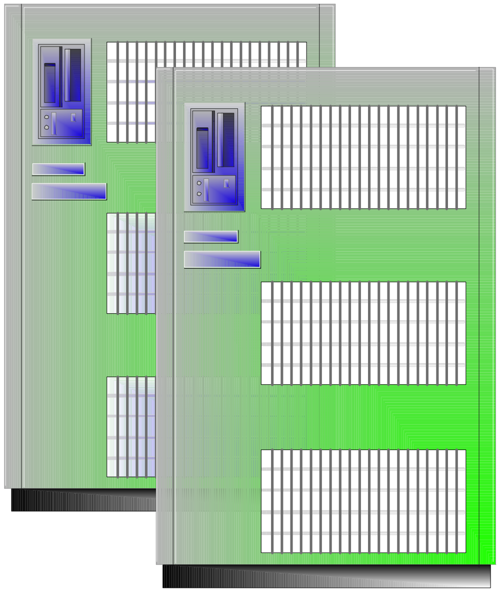
- ***Mandated by:*** **TCO, Availability and Ease of Use**
- ***Enabled by:*** **increases in processor speed and disk areal density**

Almaden
Research Center

# The Move Towards Modularity

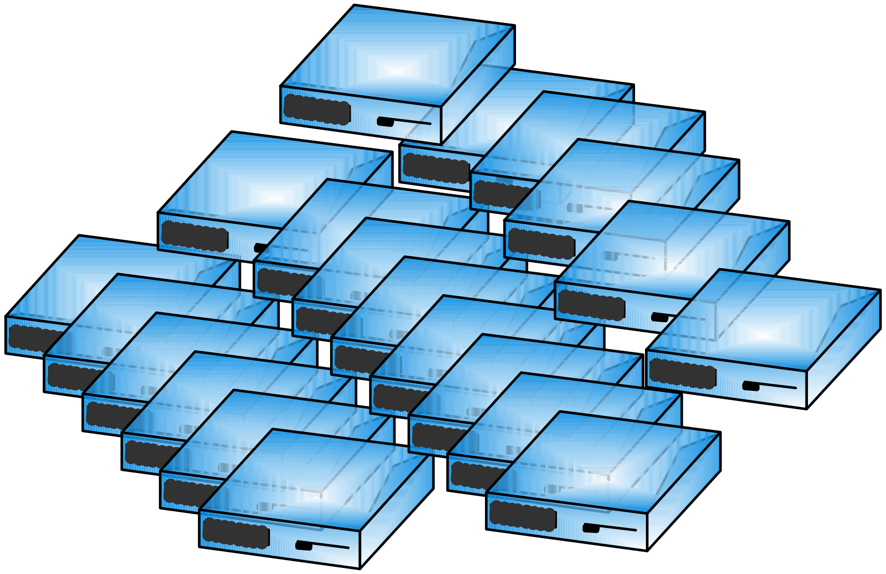
## Monolithic

- Scaling is extremely coarse
- High Management costs
- High entry cost
- Very robust components
- Failure disruptions can be major
- Failed components repaired



## Modular

- Scaling is fine grain
- Low Management costs
- Low entry cost
- Moderately robust components
- Failure disruptions are small
- Failed components not repaired



Almaden
Research Center

# Fail-in-place

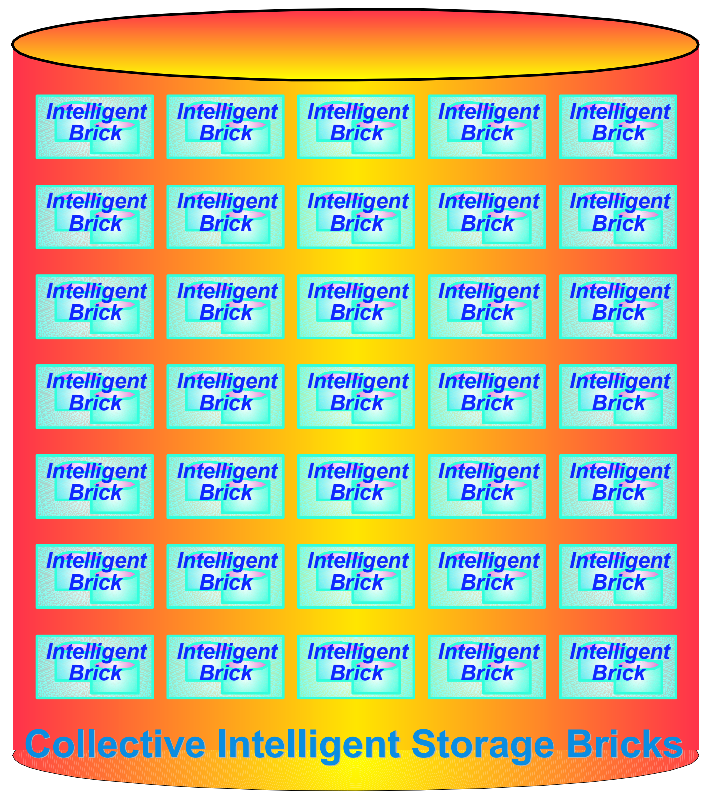
## Goal is to reduce cost by increasing availability

- Can service actions be minimized or even eliminated?
  - Many service actions result from previous service actions

- Unavailability = $\dfrac{MTTR}{MTTF}$

- To achieve better availability:

↑ MTTF or ↓ MTTR




## Collective Intelligent Storage Bricks

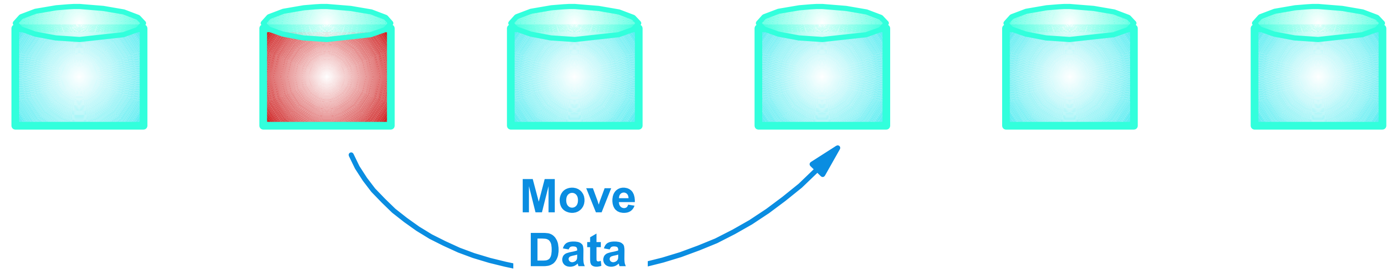
- Overprovision the system
  - Seal the bricks
- Reliability Increases by...
  - Improved sparing
  - High levels of redundancy

http://www.almaden.ibm.com/cs/storagesystems/CIB

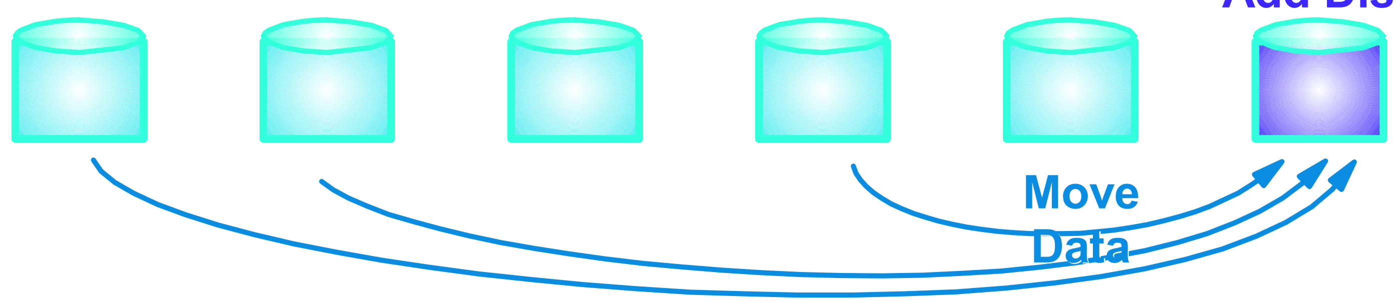
Almaden
Research Center

# Self-Management

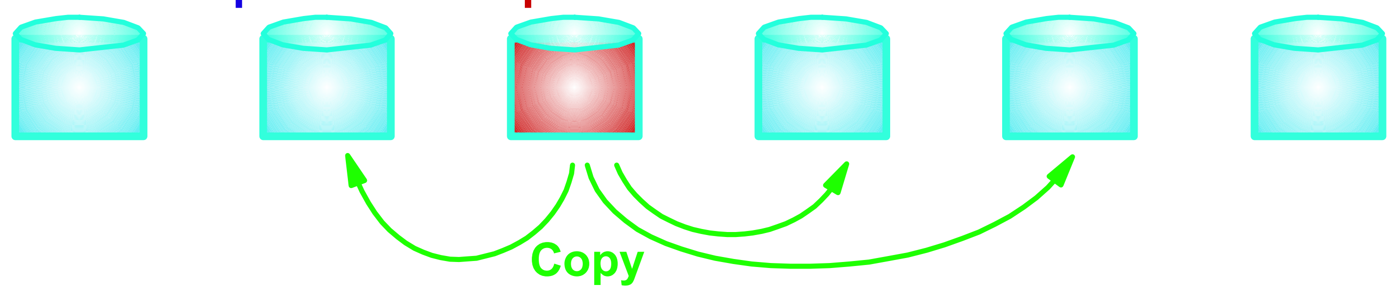
**Eliminate Hot Spots**




**Add a disk, move data and balance**




**Proactive copies for hot spot elimination**

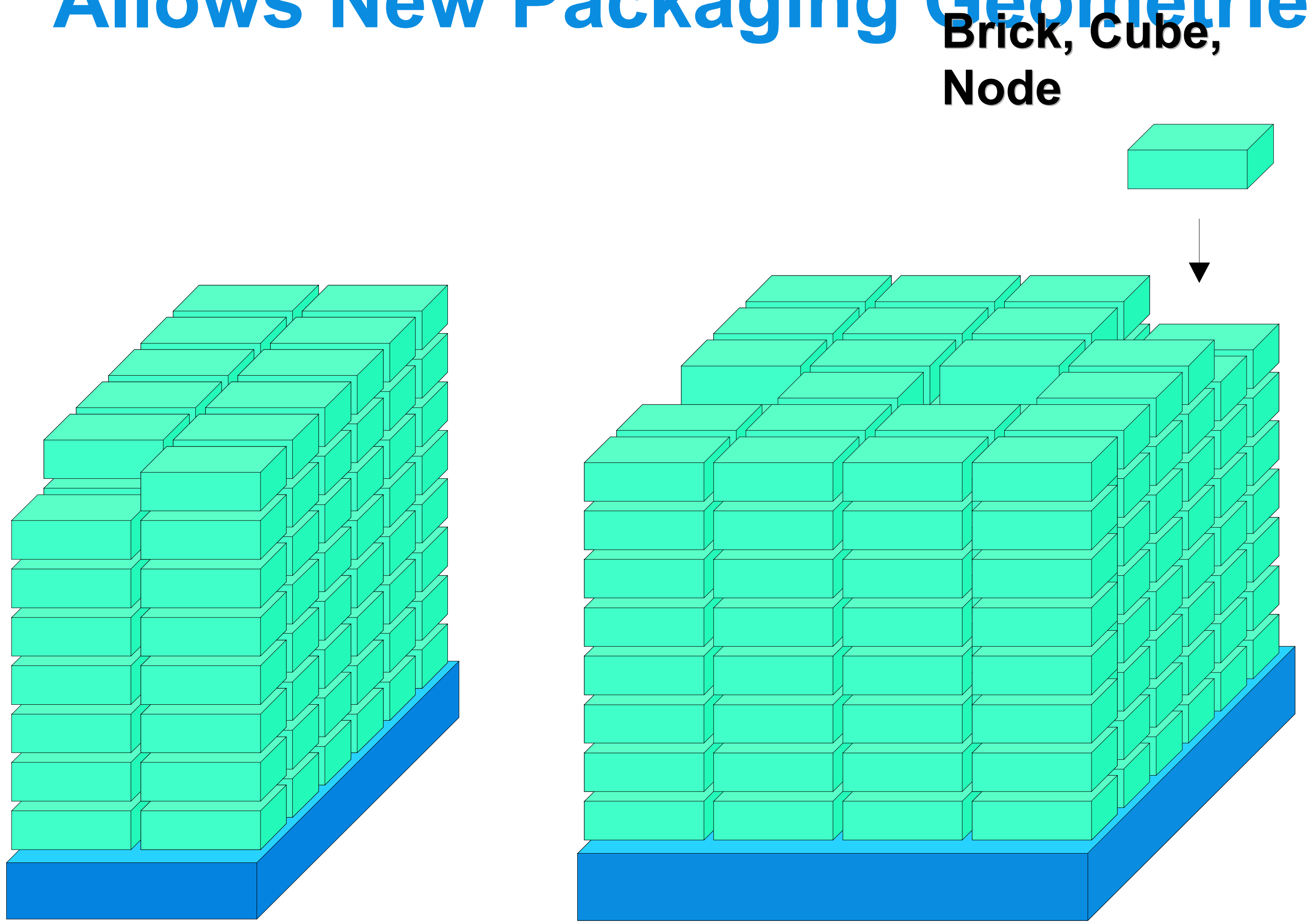



**Automatic Data Recovery**

- Traditional RAID functions (parity, mirror, etc. ...)
- Copies can be used for higher levels of reducndency

# Fail-in-place

## Allows New Packaging Geometries

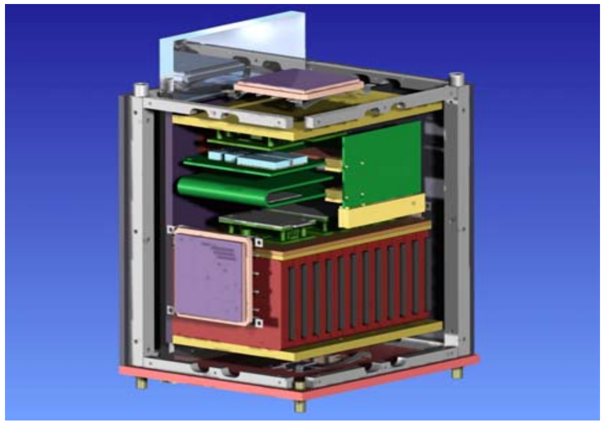



Various "Ice Cube" shapes

W.W. Wilcke, 'Comp. Arch. Trends for the next Ten Years' 25th ACSC, Jan. 28-Feb. 1, 2002, Melbourne, Australia

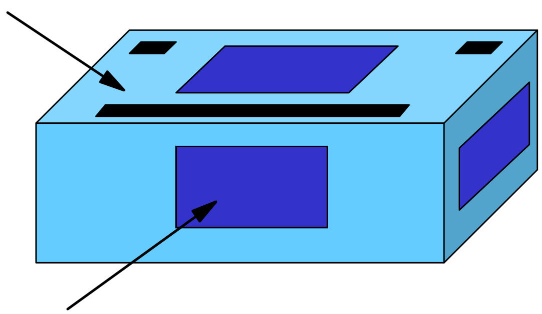
Almaden
Research Center

# IceCube Assembly

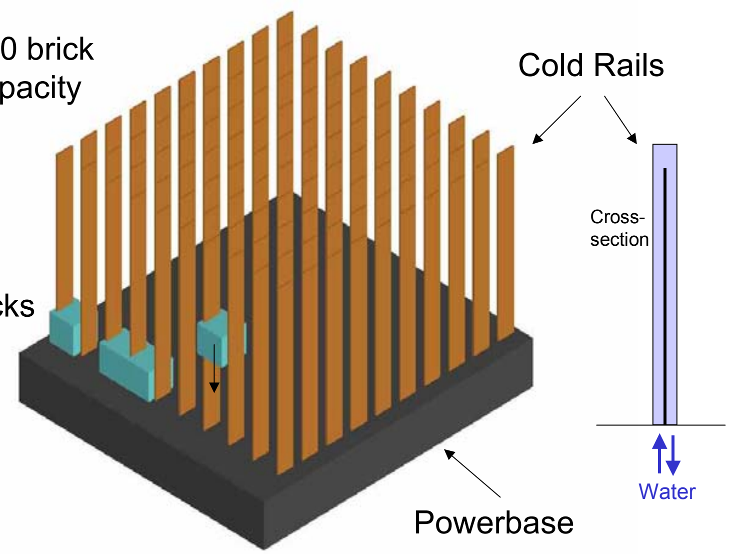


## Ice Cube Prototype  Brick

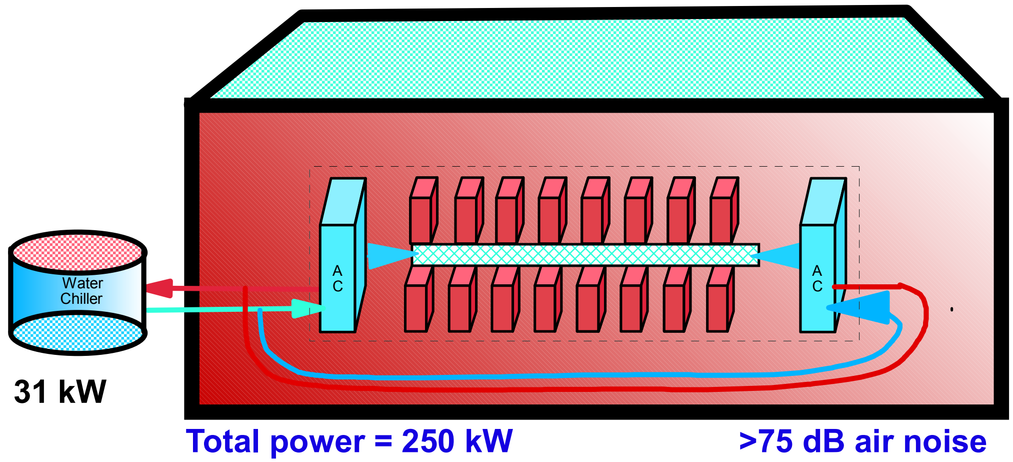
- SpecInt2000: 633
- Watt: 200
- Size: 20 cm = 7.87”



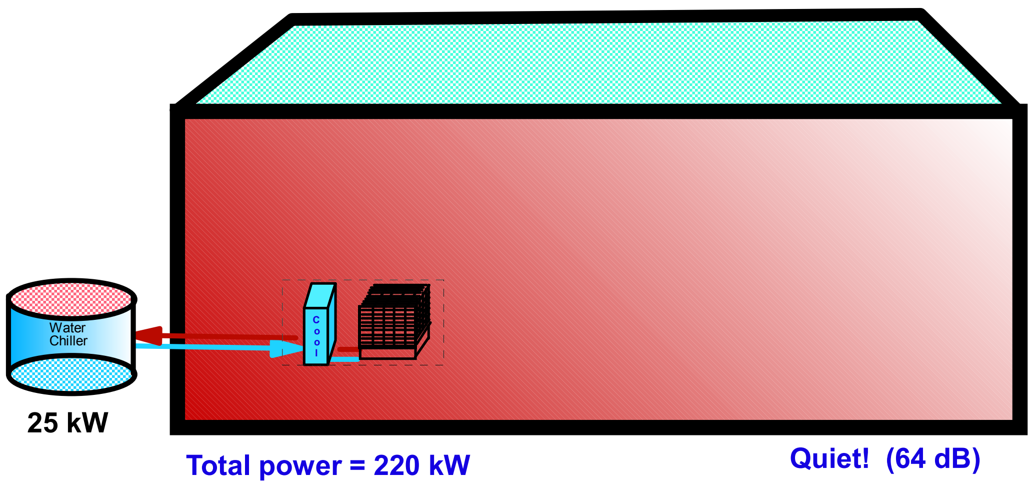

## 700 brick capacity



# Collective Intelligent Bricks
# IceCube ⇒ 1-PB




- 32 Racks
- 640 CIB
- 8 240-GB 3.5" Disks per CIB
- 275 W per CIB
- 5.5 kW per Rack




- 640 CIB
- 8 240-GB 3.5" Disks per CIB
- 275 W per CIB

http://www.almaden.ibm.com/cs/storagesystems/IceCube


# Bandwidth and Storage Virtualization

## Example 10x10x10 IceCube ....

- Few Petabyte capacity
- Bisectional Bandwidth 6000 Gbits/s in each dimension
- External Bandwidth 4000 Gbits/s
  - Ports on four vertical walls
- Latency 130 nanoseconds per hop (only!)

## Huge Bandwidth: Storage Virtualization (SV) becomes very practical

- Data can be distributed in cube nearly without regard to location
- Software for SV much easier to develop
- Most tasks of storage administrator go away
  - “just add more bricks” when SV software tells him/her

## Result:

- 1 storage system administrator per 5 TB (today)

$\Rightarrow$ 1 administrator per Petabyte (in the future)

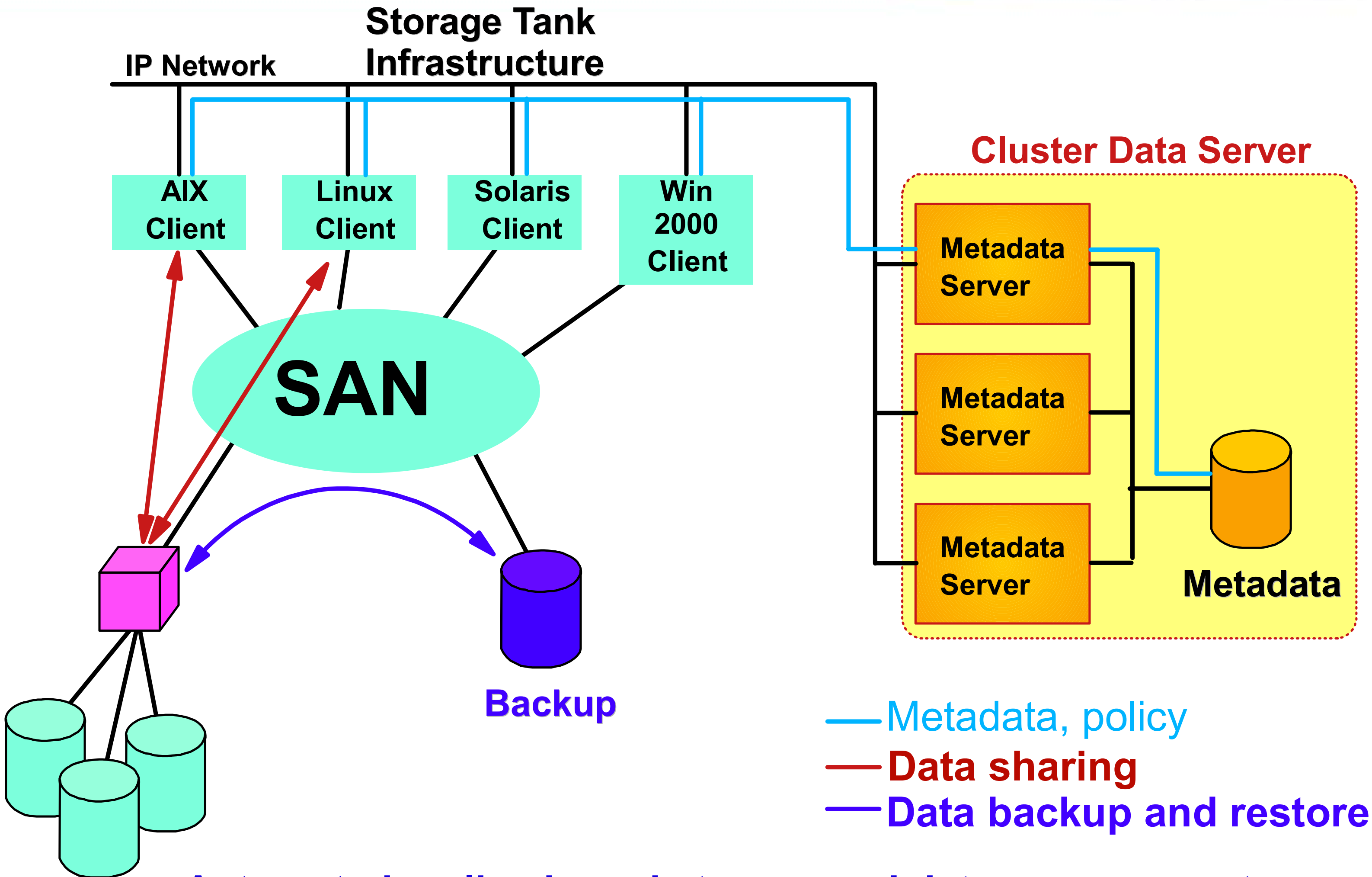

# Autonomic Computing Evolution

| Levels of Autonomic Computing Sophistication | Well known examples | Current Research and Product Directions | Future Goal |
|---|---|---|---|
| Serving the World (people, business processes) | SMS | Storage Tank | |
| Heterogeneous Components Interacting | SNMP | | |
| Homogeneous Components Interacting | Adaptive network routing, Network congestion control<br>High availability clustering | Collective Intelligence Storage Bricks<br>Oceano | New packaging concepts for storage<br>Subscription computing |
| Components | ESS<br>RAID<br>DB Optimizer<br>Virus Management | Regatta self-healing, LPAR<br>SMART/LEO<br>Software Rejuvenation | More of the same and better |

**Future Innovations**

Almaden
Research Center

# Policy Managed Storage: Storage Tank




- Automated, policy-based storage and data management
- High performance, multi-platform file sharing

**See Work-in-Progress talk this evening for more information on Storage Tank**

www.almaden.ibm.com/cs/storagesystems/stortank


# Storage Devices and Systems: Key Drivers of the IT Industry

## Storage Devices Challenges

- Fundamental limits to be overcome
- Prospects for new technology, materials, etc...

## Storage Systems Challenges

- Software for implementing policies automatically
- Intelligent modular hardware
- TCO
- Reliability
- User Experience

Serving business and societal needs

# *Thank you!*