# Perfect Devices:
# The Amazing Endurance of Hard Disk Drives

Giora J. Tarnopolsky

TARNOTEK & INSIC -

Information Storage Industry Consortium

**www.tarnotek.com**
**gjtarno@tarnotek.com**
**www.insic.org**

NASA/IEEE MSST 2004
12th NASA Goddard/21st IEEE Conference on
Mass Storage Systems & Technologies
The Inn and Conference Center
University of Maryland University College
Adelphi MD USA
April 13-16, 2004

# Outline

- Perfect Inventions
- Hard Disk Drives & other consumer products
- Hard Disk Drives: Developments 1990 - 2004
  - Marketplace
    - How the technology advances have affected the product offerings
  - Technology
    - How market opportunities propelled basic research forward
- Disk Drives at the Boundaries
- INSIC and Data Storage Systems Research
- Closing Remarks: Hard Disk Drive Endurance

# PERFECT INVENTIONS

# Nearly Perfect Inventions

- Certain inventions are created "perfect:"

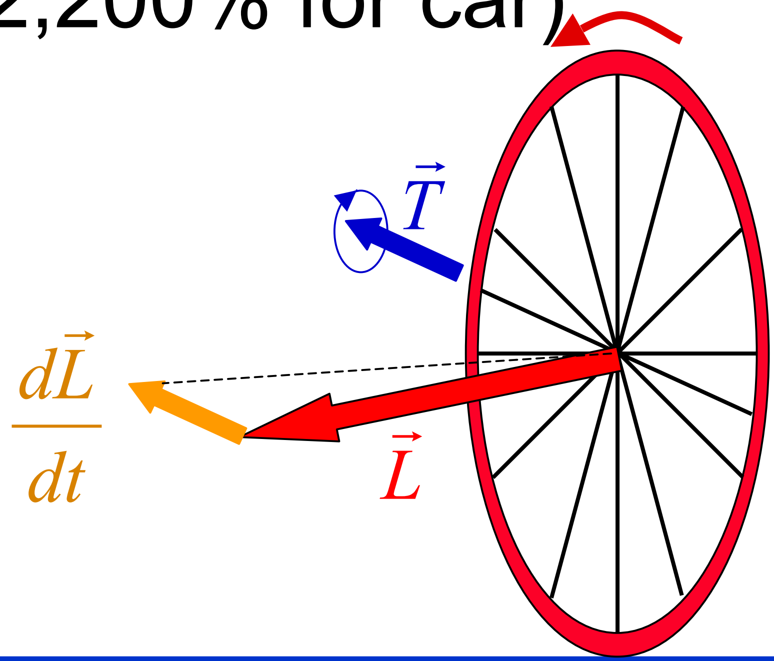
  their operation relies on a fundamental principle that cannot be improved, or does not merit improvement

  - This assures their endurance …
  - … and defines their domain of development, the limits of applicability of the invention

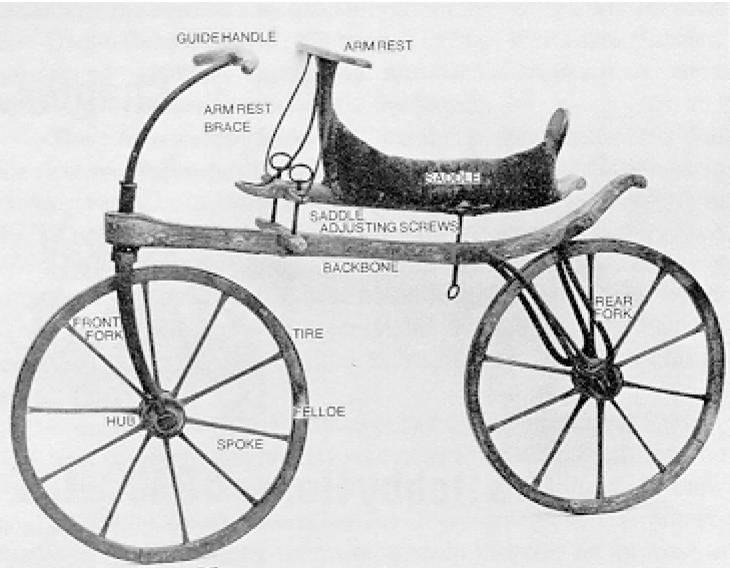
- Examples of perfect inventions are the bicycle, the umbrella, the book, and the disk drive

# Bicycle

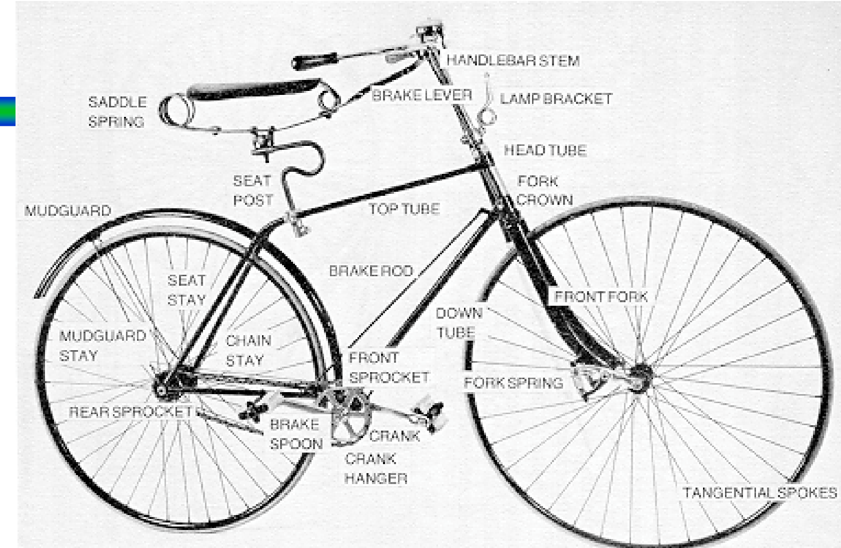
- Gyroscope effect assures stability of the rider
  - Under torque $T$, the bike turns but does not fall
- Low ratio of vehicle mass to rider mass
  - ~ 15 % (as compared to ~2,200% for car)
- Efficient
- Rugged
- Mass-produced
- Affordable



# Development




Draisienne **1817**




Columbia Light Roadster **1892**

Santa Cruz **2002**

- $dL/dt = T$ cannot be improved upon

# Disk Drive

- Magnetic hysteresis
- 2-D travel with only one linear motion
- High volumetric density
- Random access
- Mass-produced
- Non-volatile
- Affordable
- Rugged

➔ Few-hundred $/box

➔ No vibration isolation

➔ no T stabilization

➔ These properties define drives

# Development

IBM RAMAC
1953

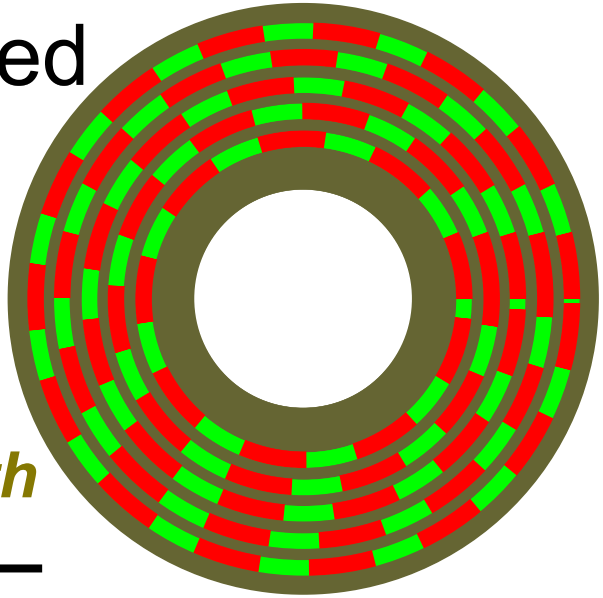
Seagate's ST506 1979

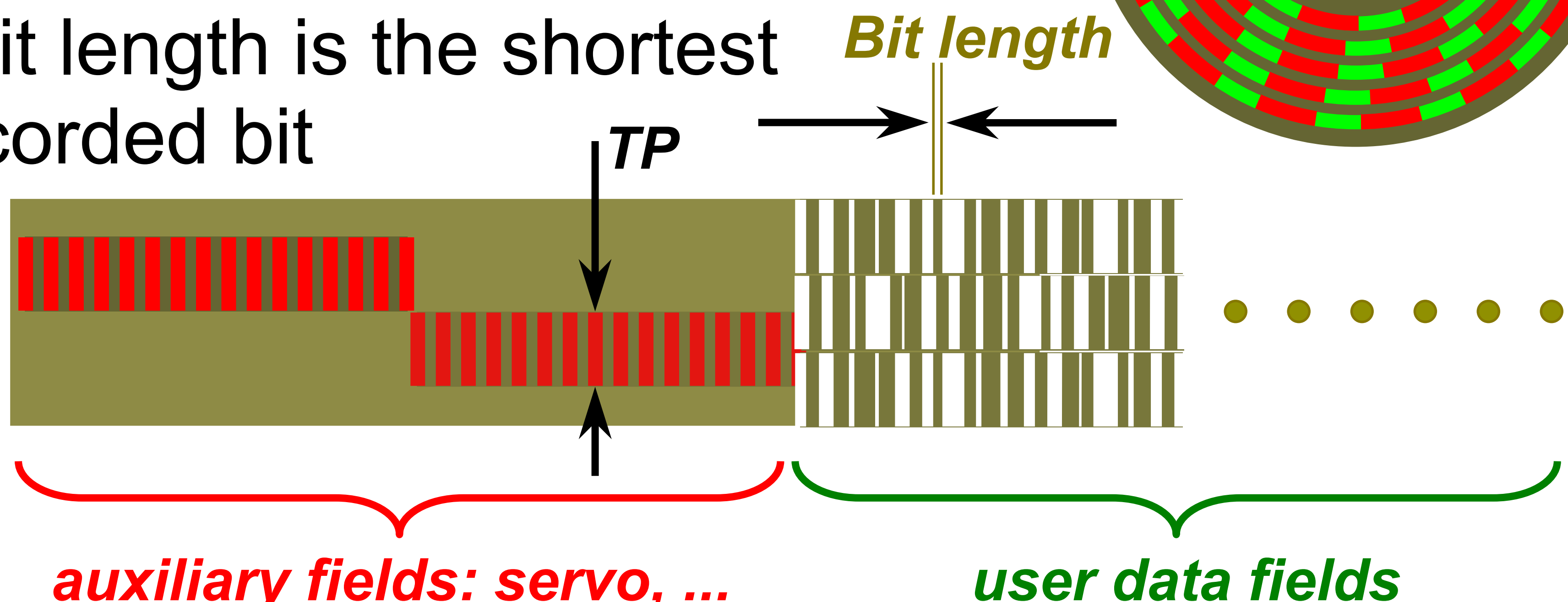
ST 'Cuda V
2002

# Hard Disk Drives

# Hard Disk Drives: Nanotechnology

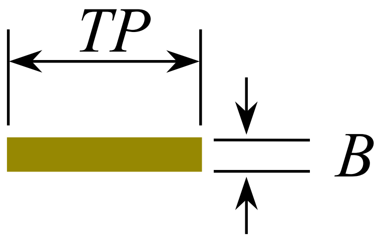
- In a disk drive, the data is recorded in concentric tracks
- Track pitch, TP, is the distance between track centers
- Bit length is the shortest recorded bit





# Hard Disk Drives now: 60Gb/in$^2$

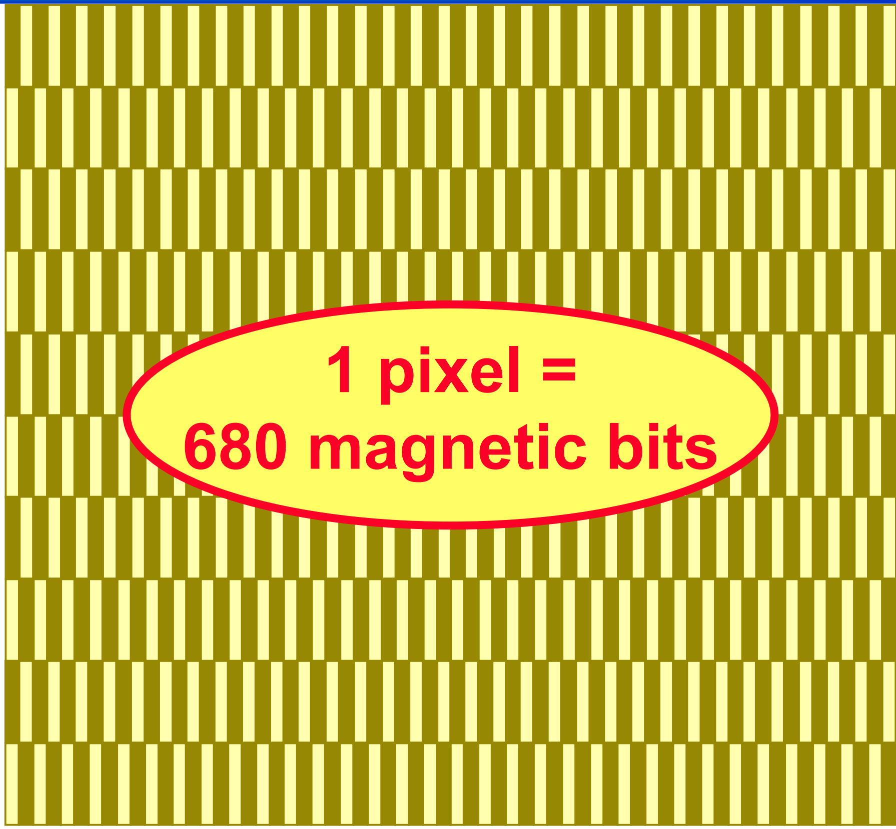
- Areal density, AD, is the number of magnetic bits per unit area
- At 60 Gb/in$^2$, the track density is 100,000 tracks per inch, and the bit density along the track is 600,000 bits per inch (100 ktpi x 600 kbpi)
- TP = 10 µin = 254 nm
- B = 1.7 µin = 42 nm




- The servo tracking mechanism *1-σ* error (or misregistration) is 8.5 nm

# HDD’s & other consumer products

- **5 megapixel digital camera: pixel ~2.7 x 2.7 μm$^2$ , or *680 times less “dense”* than bit storage**



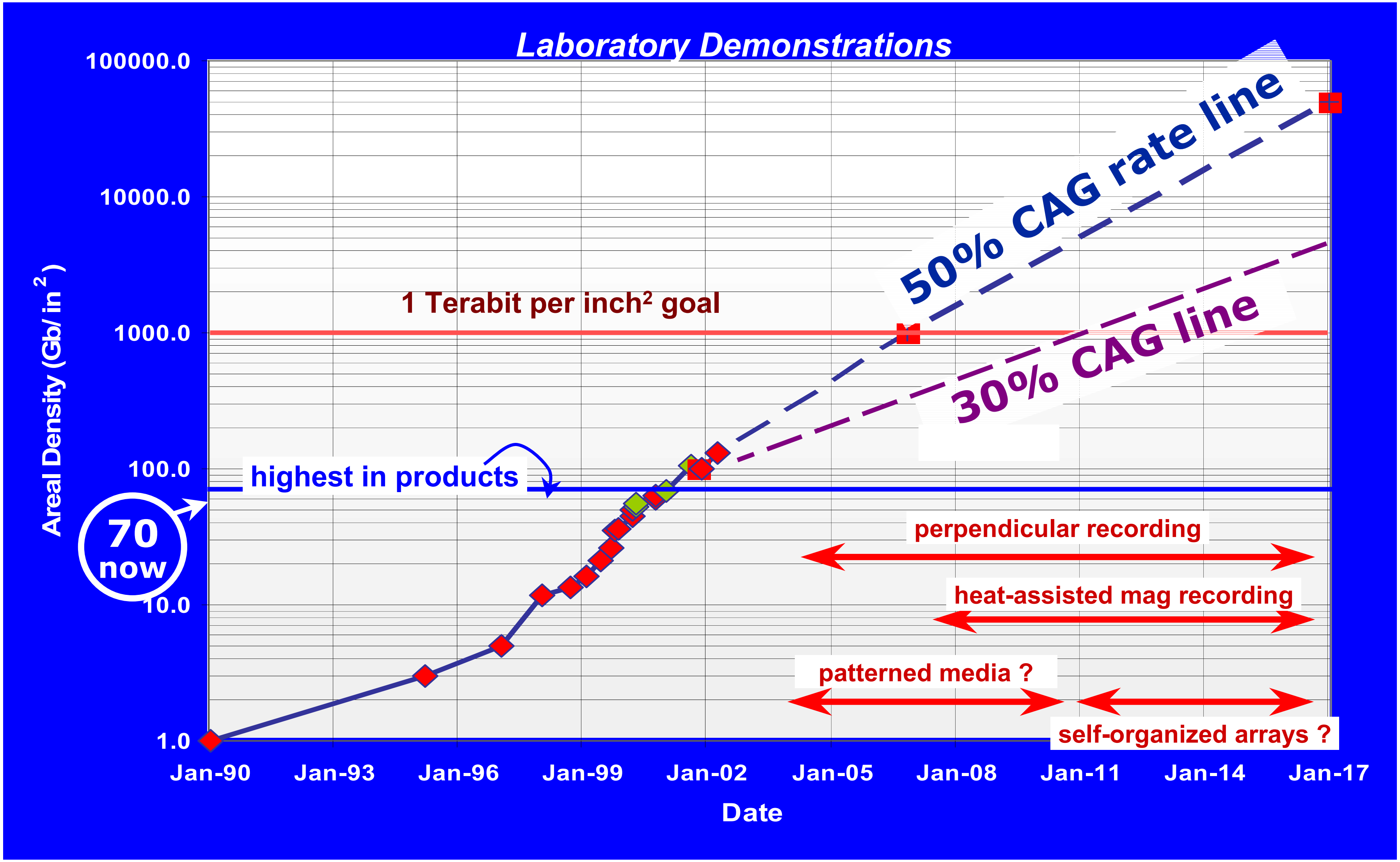

- **The focusing accuracy of the lens-to-CCD distance is ±7.5 μm, or *882 times lower precision* than HDD tracking**

- **5 megapixel camera, $379. 200 GB drive, $100. ~ x1,000 mechanism, 1/4 price.**

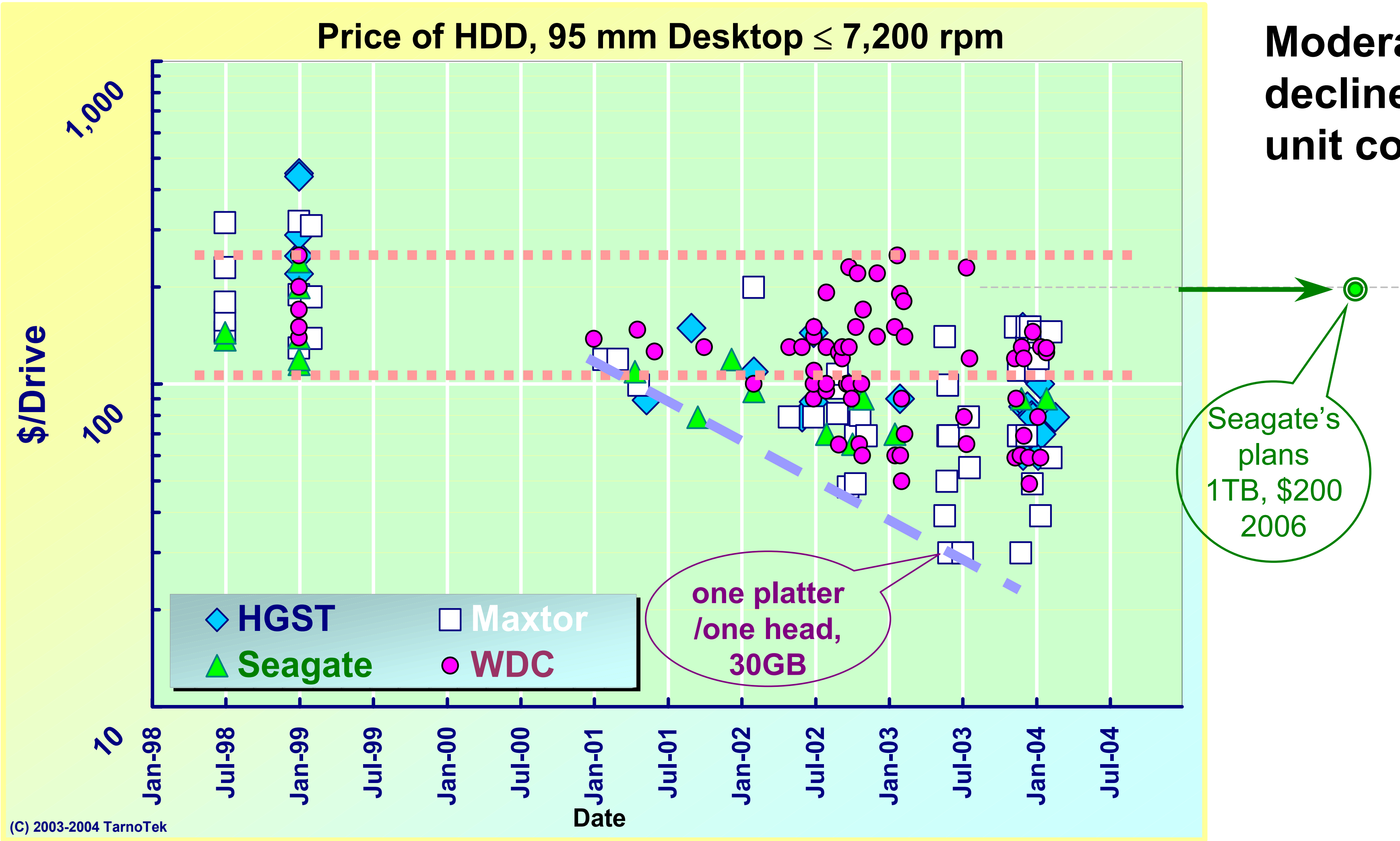
# Technology Advances Rapidly Deployed into Products

# The Future of Hard Disk Drive Technology

## Lab Demos: Possible HDD Areal Density Progression

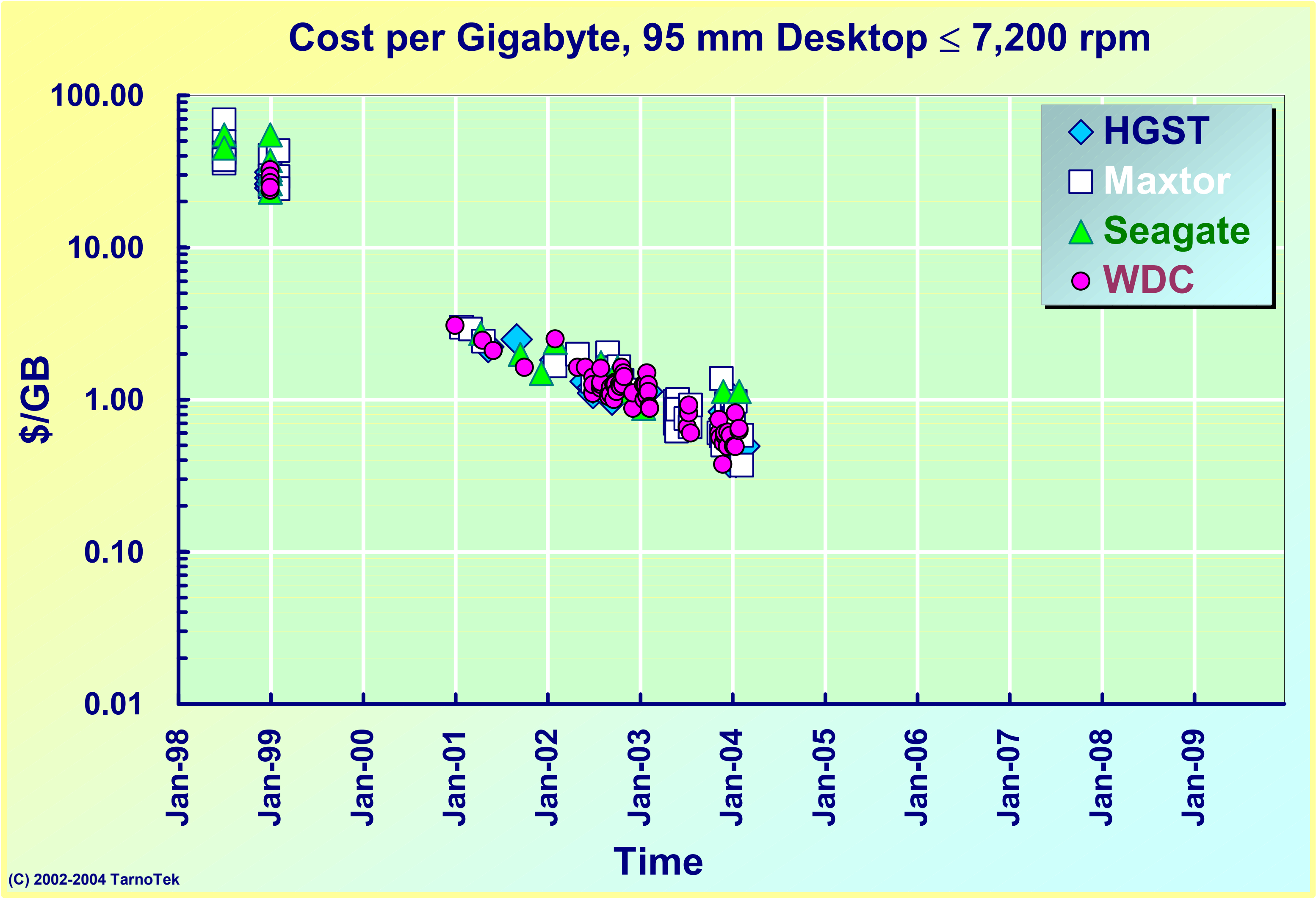


# $ of HD Drive ≥ $ Components

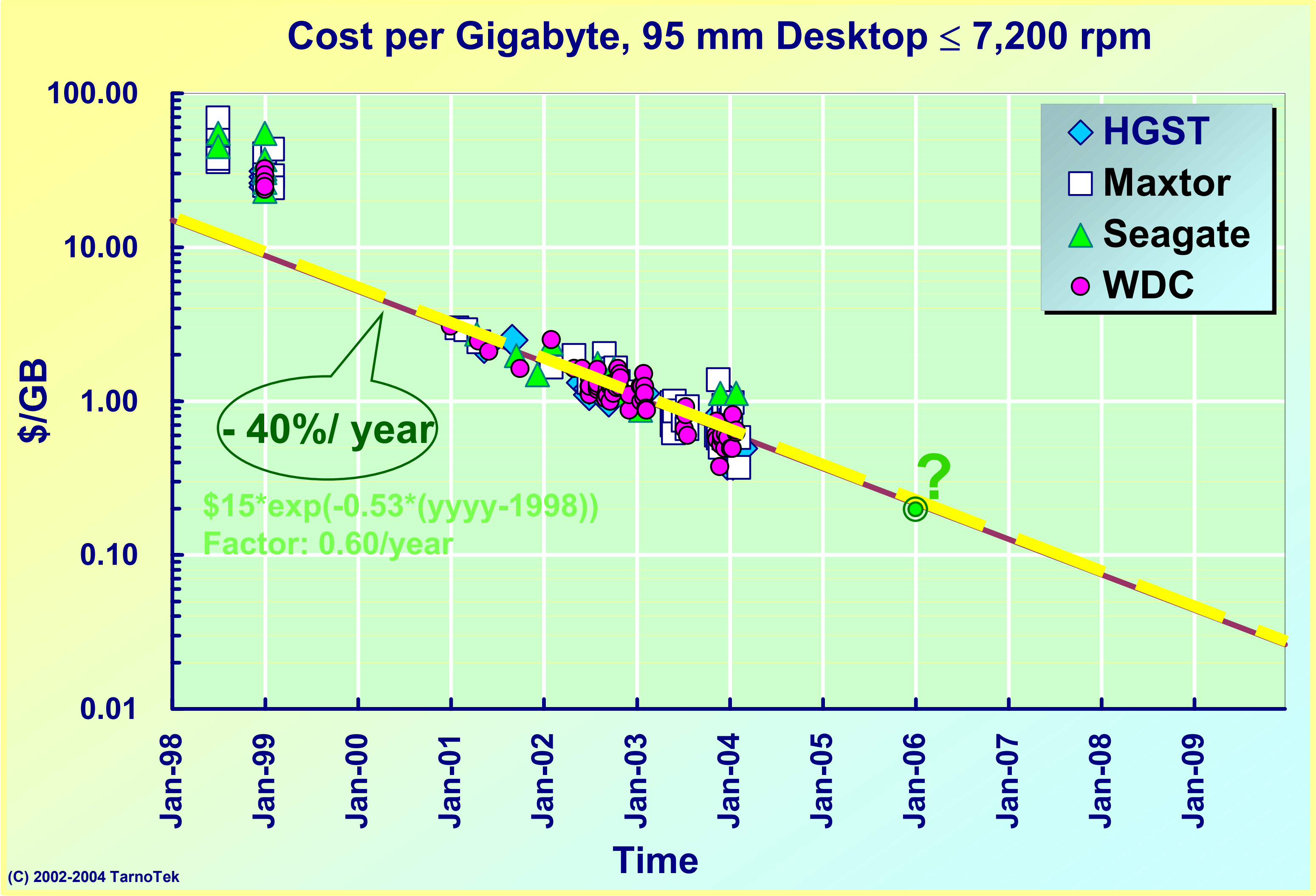



**Moderate decline in unit cost**

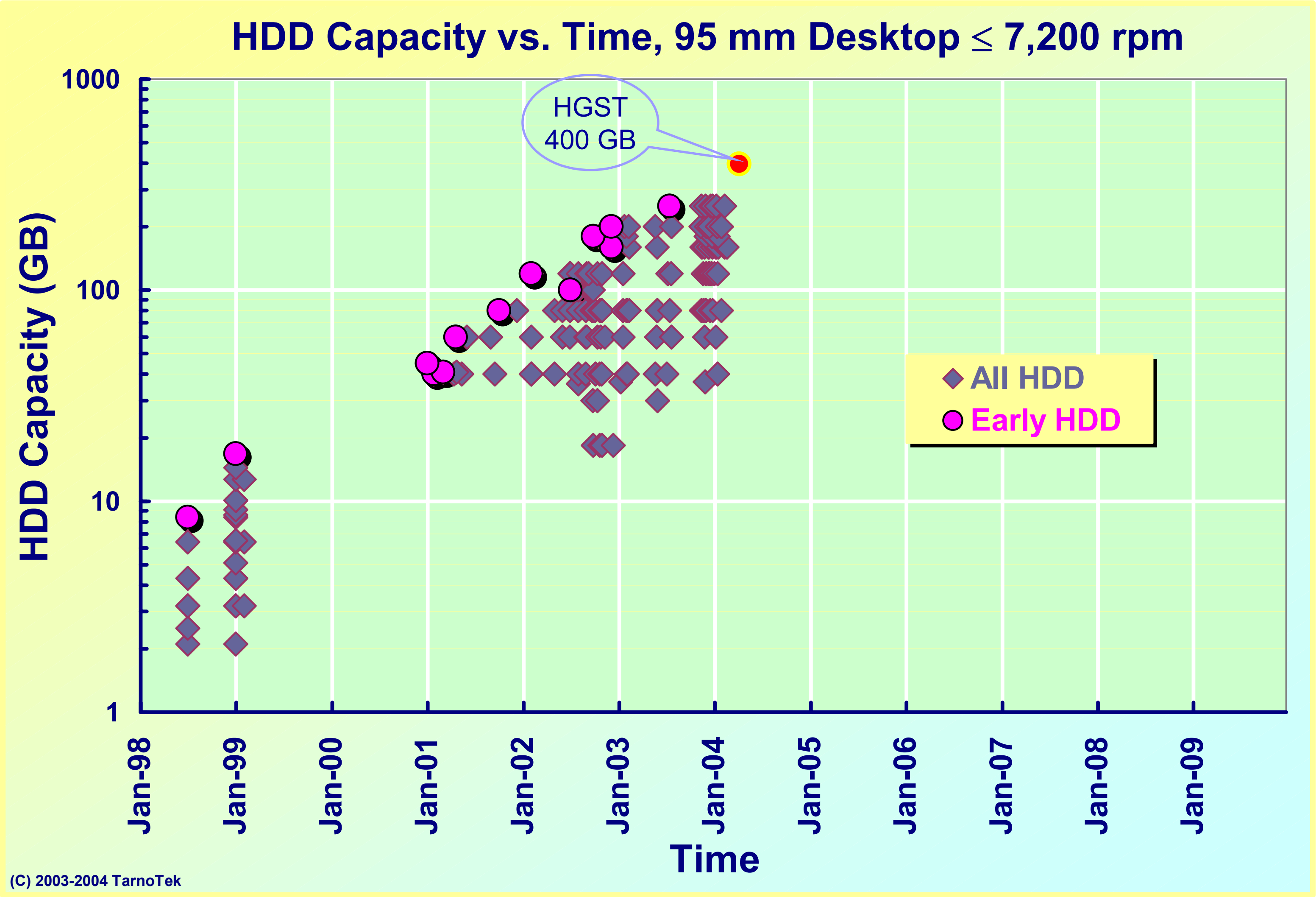
# Precipitous decline in $/GB



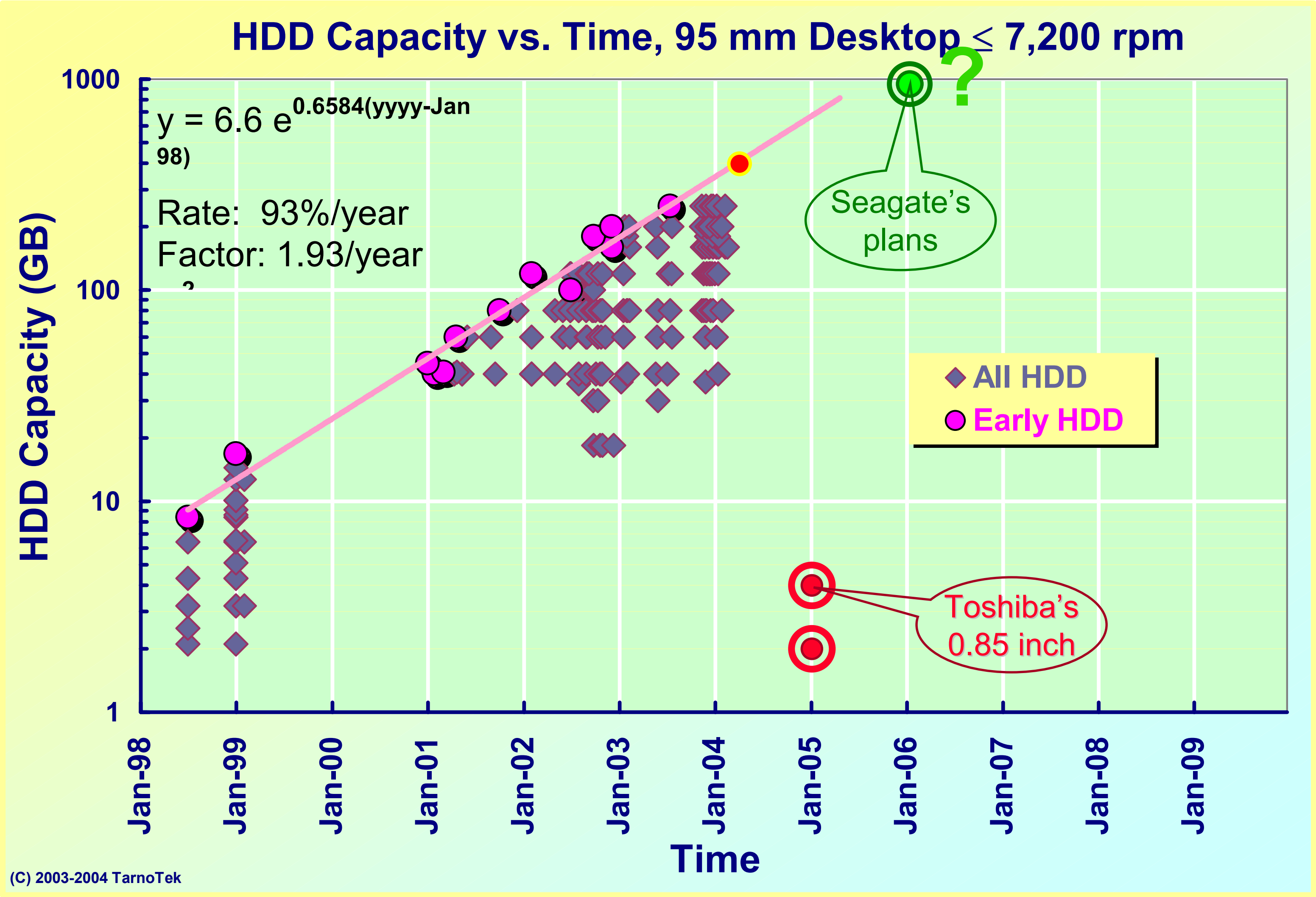
# Precipitous decline in $/GB



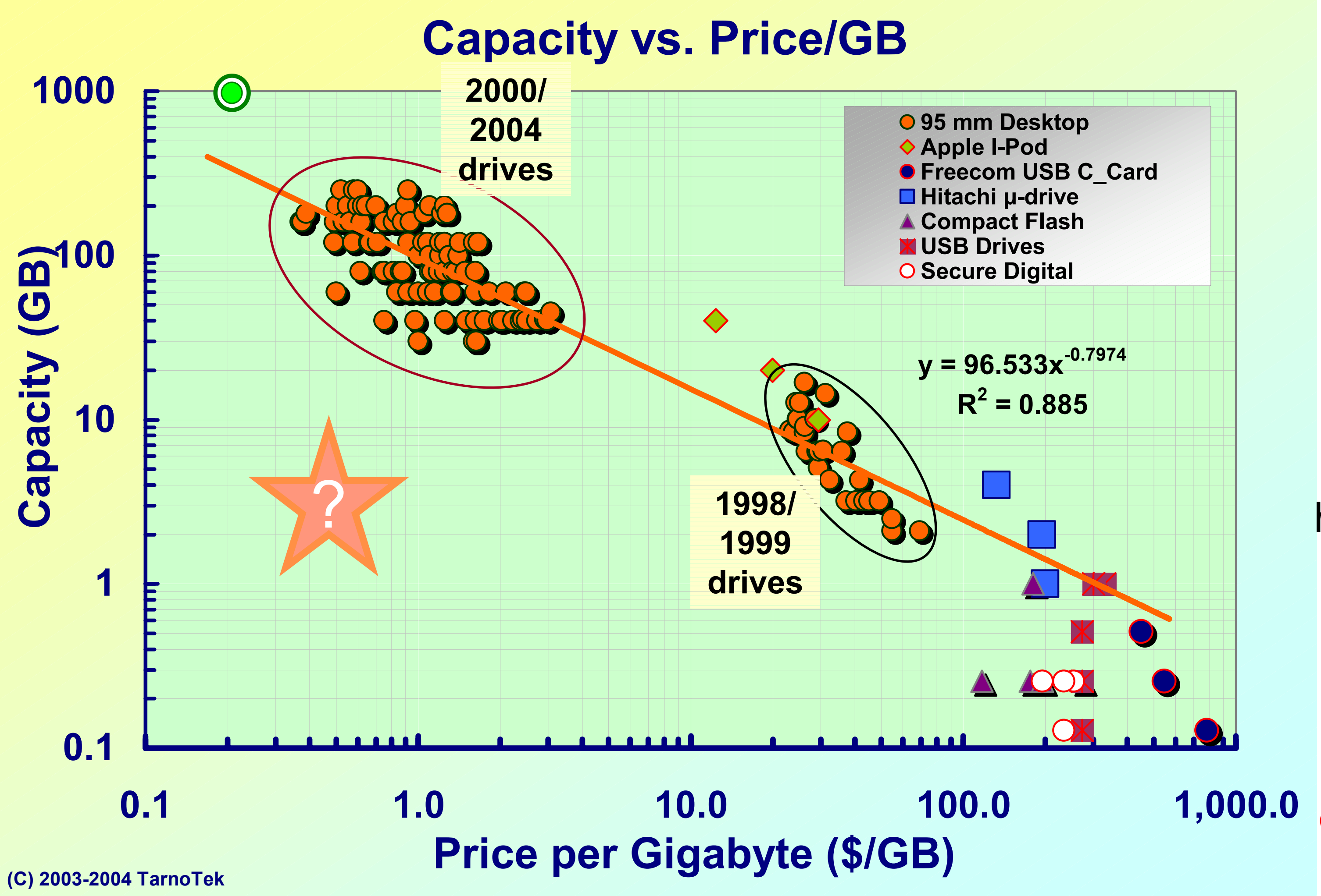
# "Box" Capacity Growth



# Capacity Growth: Sustainable?



# Product Capacity vs. Normalized Cost




- Capacity inversely proportional to $/GB
- AD growth invested almost exclusively in capacity growth
- Increased AD & miniaturization have not reduced cost at desirable capacities
- No 10 GB, $5 HDD products
- **Beyond some capacity, capacity itself is not a customer need.**

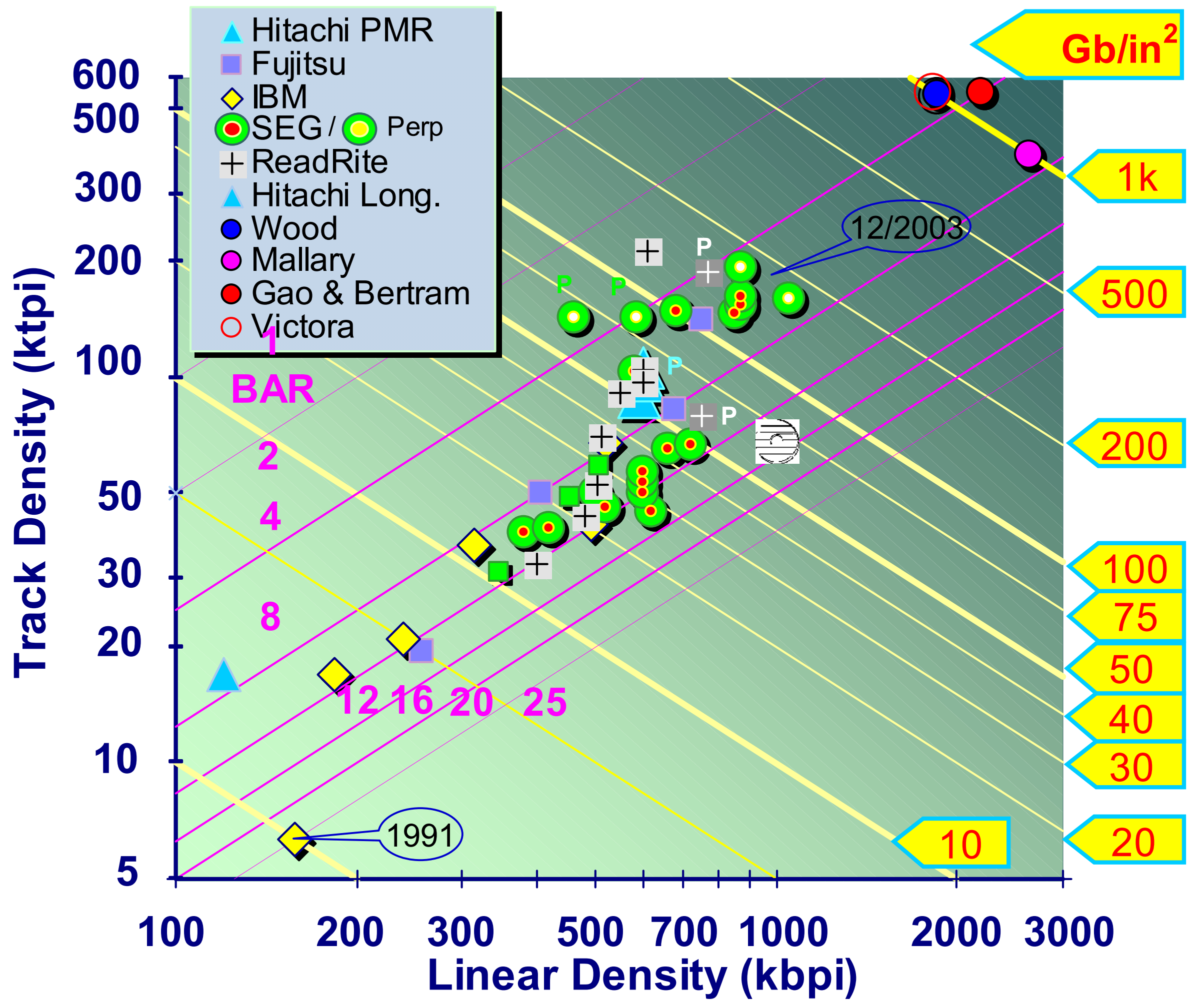
# Magnetic Areal Density Progression

# Areal Density Growth

- Laboratory magnetic areal density has grown from 1 $Gb/in^2$ in 1990 to ~170 $Gb/in^2$ by December 2003
- It is 60 to 70 $Gb/in^2$ in current products
- Feasibility assessments of 1 $Tb/in^2$ under intense scrutiny, stated INSIC research goal
- 10, 50 $Tb/in^2$ have been suggested for HAMR, Heat Assisted Magnetic Recording

***HDD future predicated on continuing growth of the areal density. Or does it?***

# From 1 to 170 Gb/in²: 1990 - 2003




Superior achievement

Log- log plot of track density and linear density

Negative slope lines: constant AD

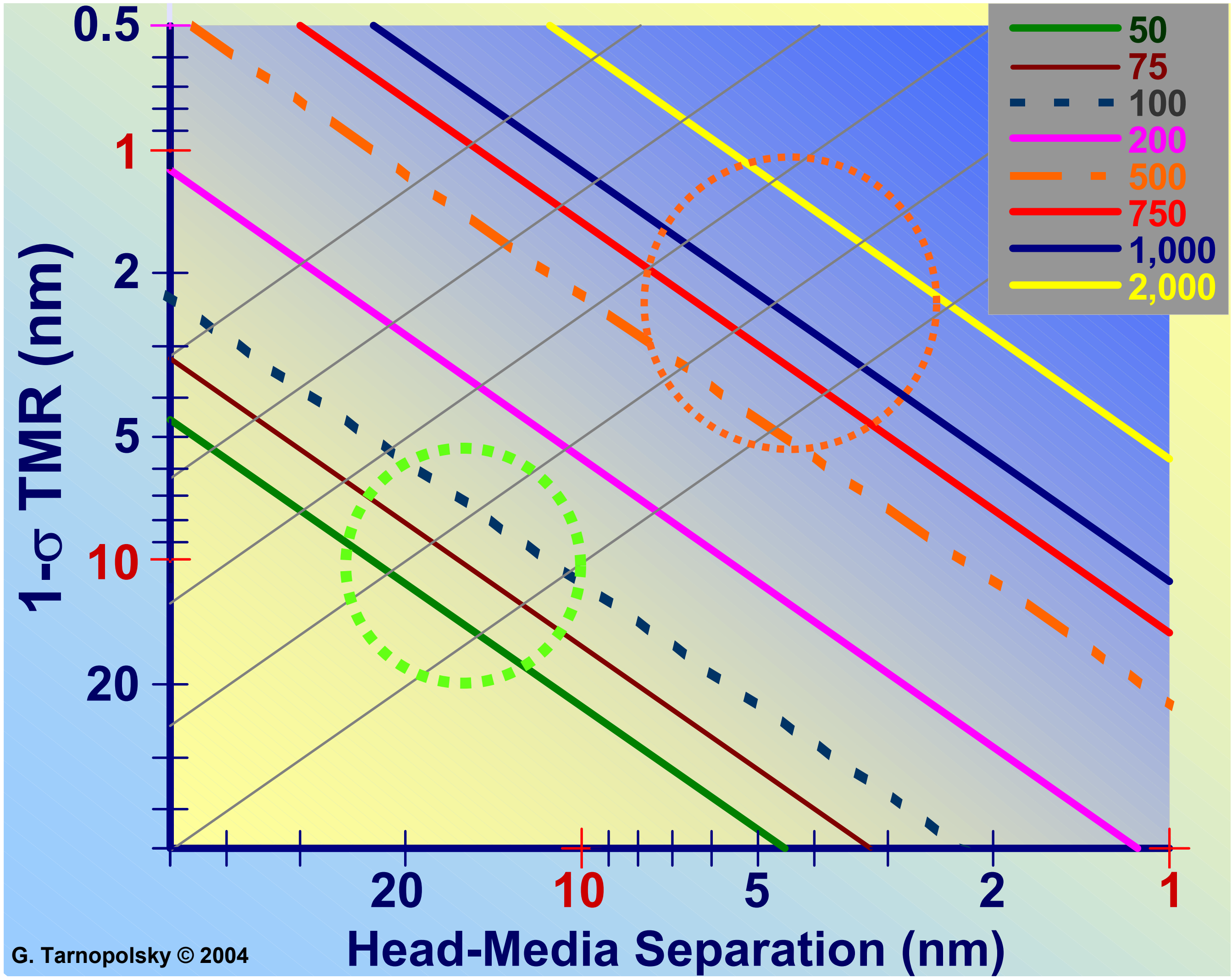
Positive slope lines: constant BAR

**Low AD demos: $10^{-9}$, much better BER than recent demos**

Since 1 Gb/in², the BAR has changed from ~ 25 to ~ 4 to 6

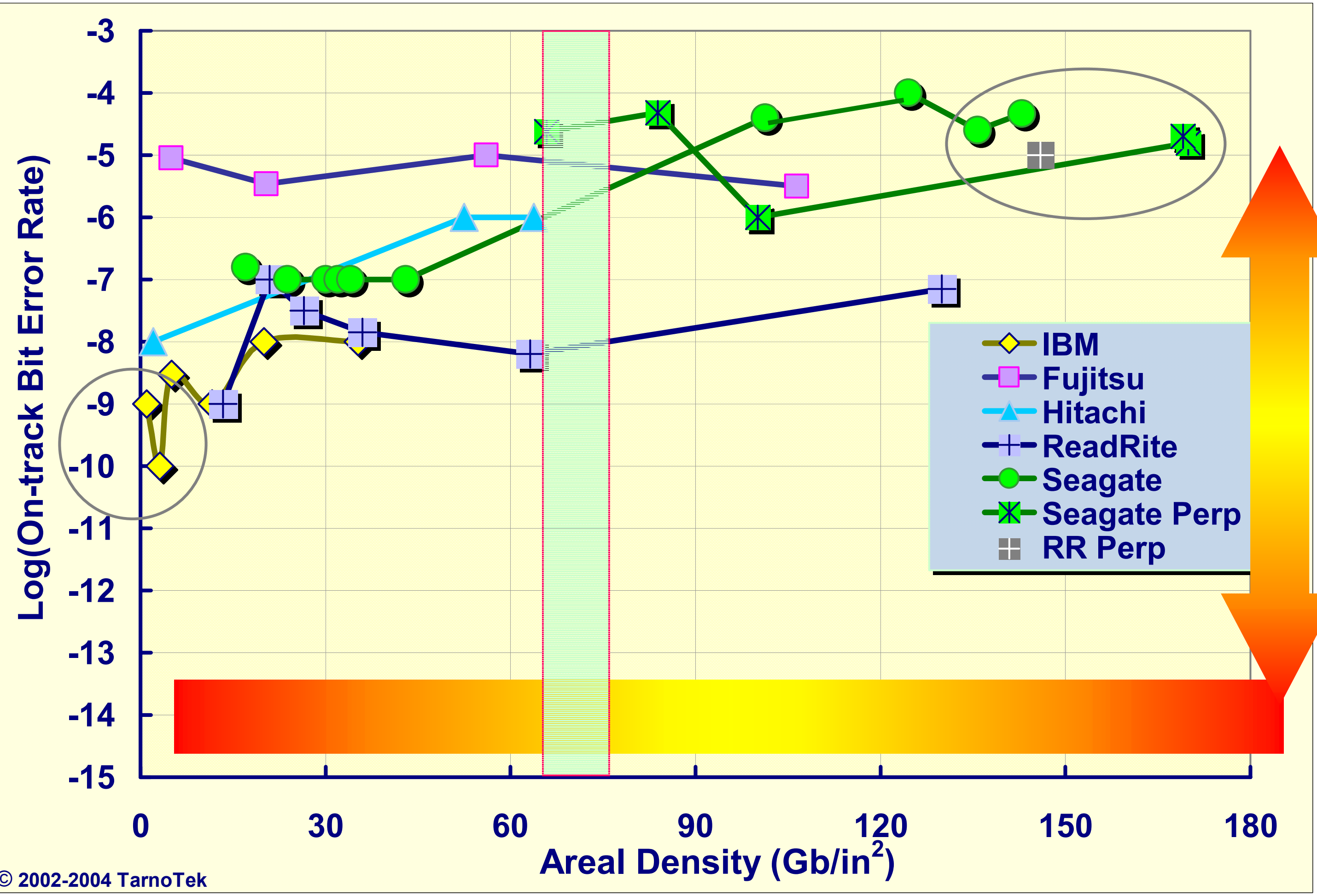
**Track density has grown faster than linear density**

# Extreme Mechanical Constraints



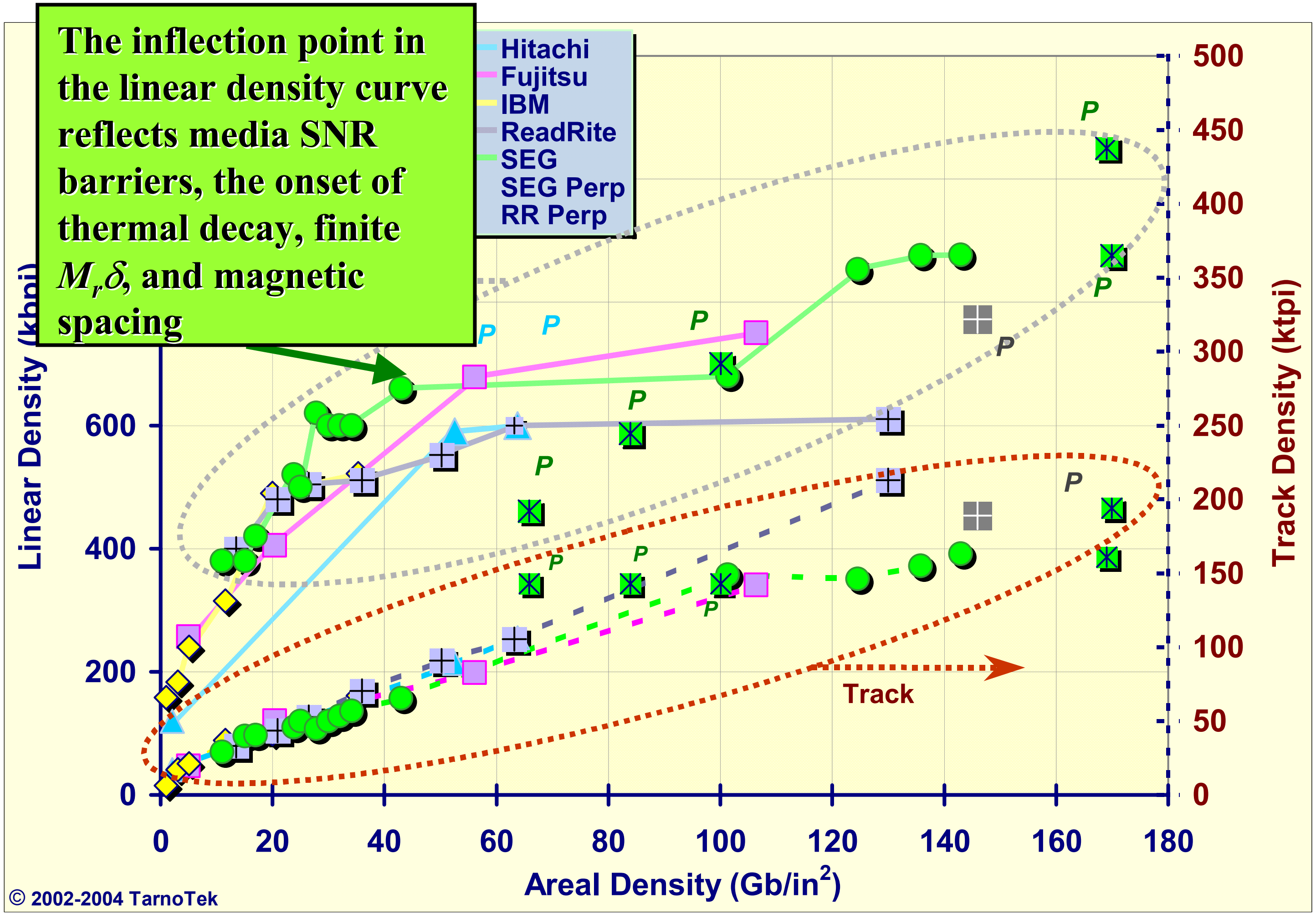

- Constant areal density curves on a TMR/HMS grid
- 3-σ TMR = TP/10
- 2π[HMS]/λ = fix
- At areal density > 500 Gb/in$^2$ ,fly height and tracking critical dimensions are one to few nanometers
- Rugged, reliable mechanism?

# A Moving BER Target



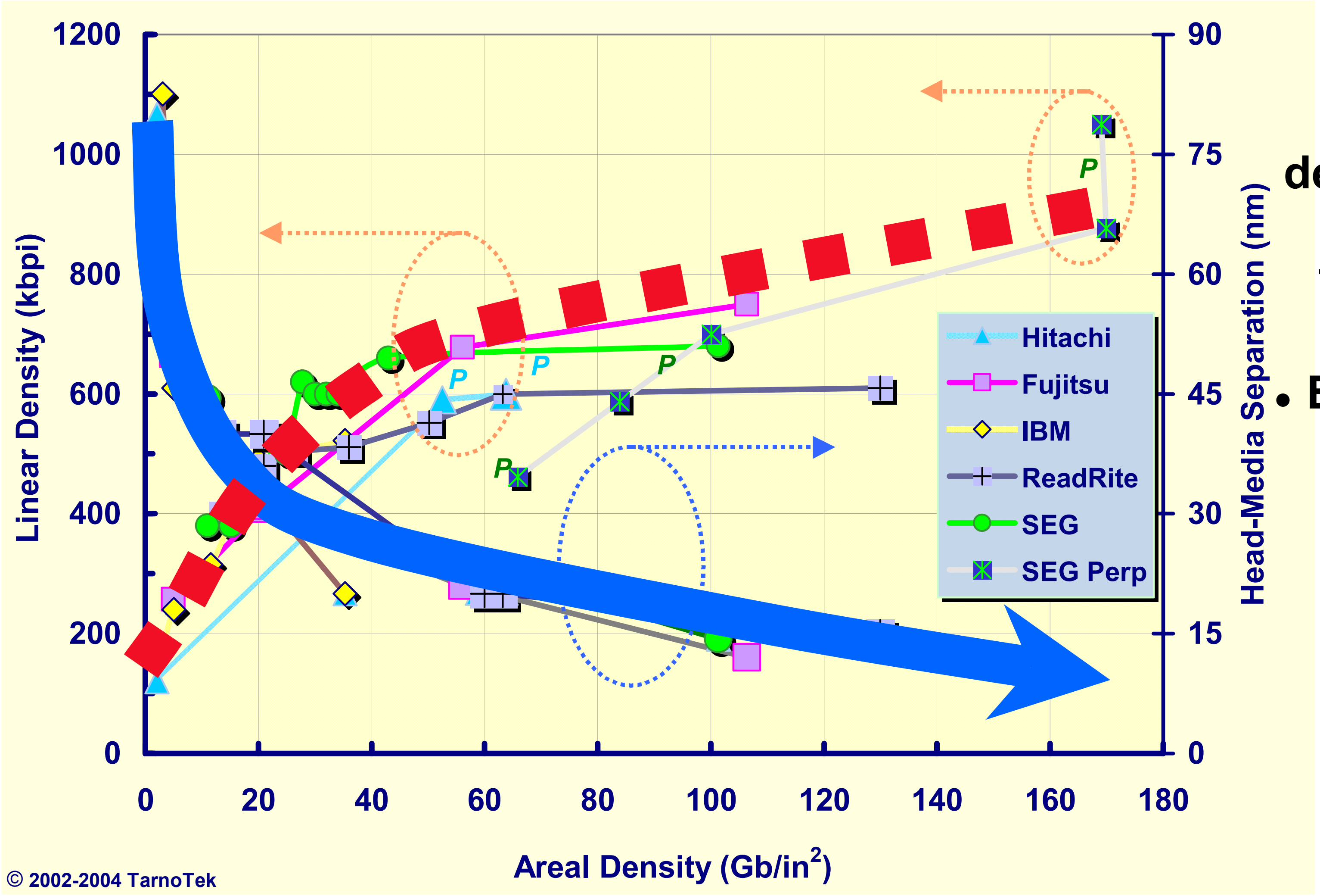

- Chart shows BER value for the experiment
- Areal density demonstrations are rigorous, comprehensive assessments of the technology
- The experiment BER has worsened with increasing areal density of demos
- The 100 ~ 200 Gb/in$^2$ regime is still R&D in 2004

# How the Areal Density was Won




- Linear and track density vs. areal density
- Track density increased by ~ 30, caused most of the areal density gain.
- Enabled by the advent of spin valve and GMR heads, advances in head fabric-ation techniques, media SNR

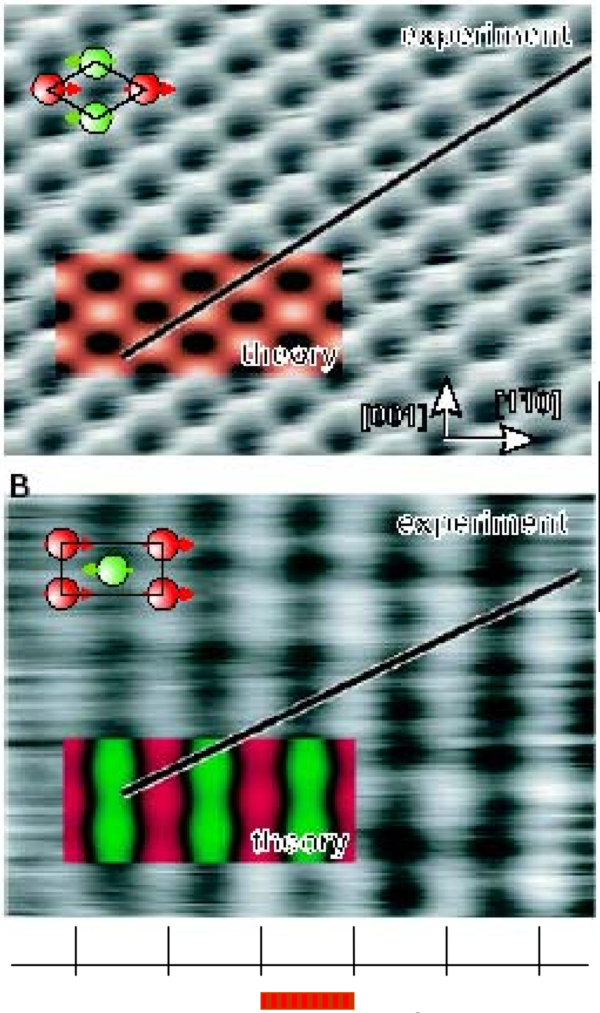
# Head-Media Separation: Nanometric




- As the recording wavelength decreased, the HMS shrunk from ~ 80 nm to ~ 10 nm
- Extraordinary efforts put in reduction of pole-tip recession, carbon overcoat thickness, flying height and lube

# Magnetic Areal Density Prospects

# 265 Tb/in² "virtually demonstrated"




4.5 ± 0.1 Å

**non-magnetic W tip**

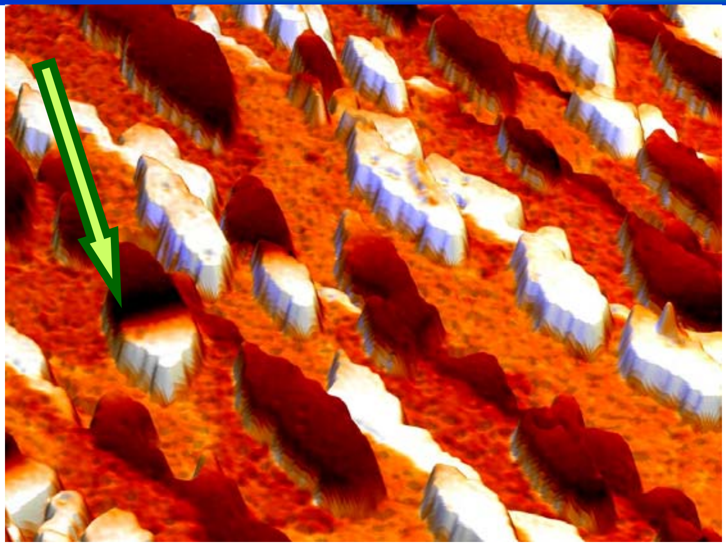
**Full image size is 2.7 x 2.2 nm²**

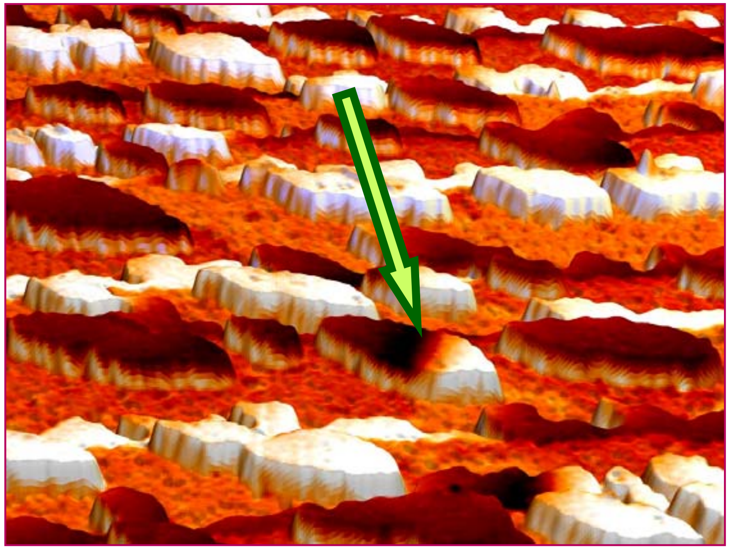
**magnetic Fe tip**

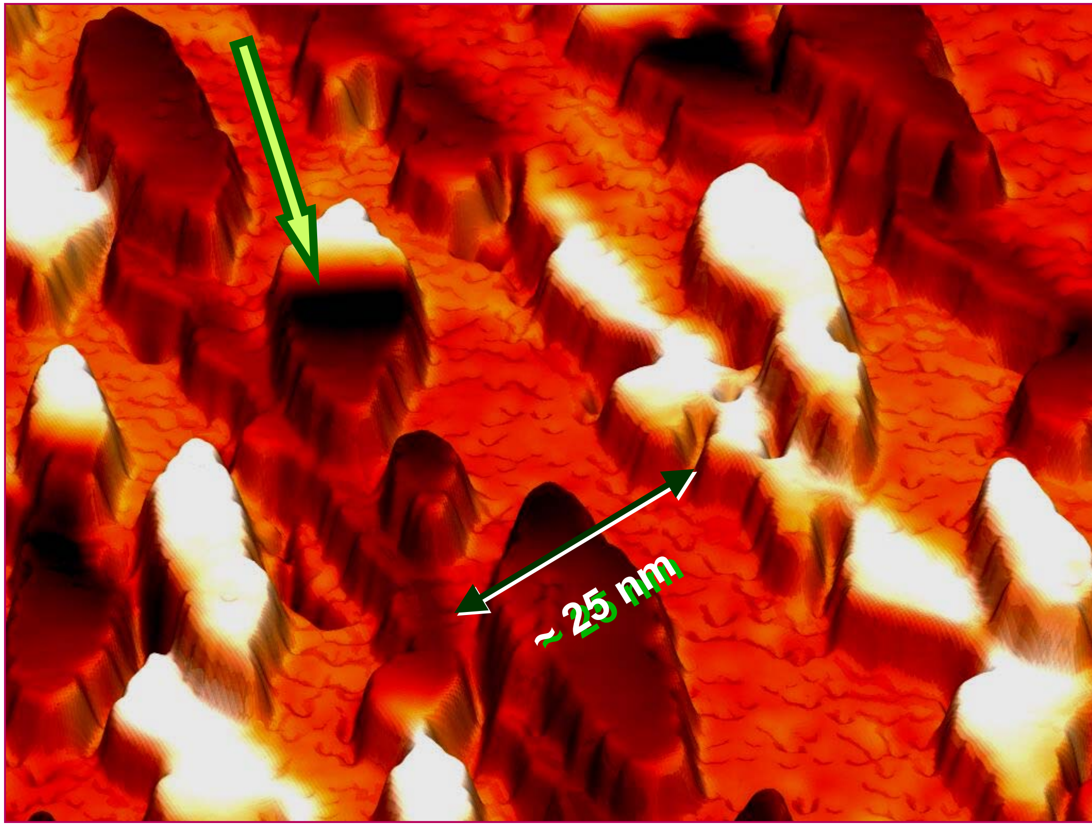
- Single atomic layer of Mn on W(110) forms a two-dimensional antiferromagnet.
- These are magnetic monolayers of chemically equivalent atoms, where adjacent atoms at nearest-neighbor sites have opposite magnetic moments.
- The ***single-atom spin has been resolved*** by Spin-Polarized Scanning Tunneling Spectroscopy
- AD ≥ 265 Tb/in² for a 12-atom bit

- **S. Heinze, M. Bode, A. Kubetzka, O. Pietzsch, X. Nie, S. Blügel, R. Wiesendanger, Science 288 (2000) 1805-1808: Real-Space Imaging of Two-Dimensional Anti-ferromagnetism on the Atomic Scale**
- **R. Wiesendanger & M. Bode, Solid State Communications 119(2001) 341-355: Nano- and atomic-scale magnetism studied by spin-polarized scanning tunneling microscopy and spectroscopy**

# Perpendicular Fe “nanobits”

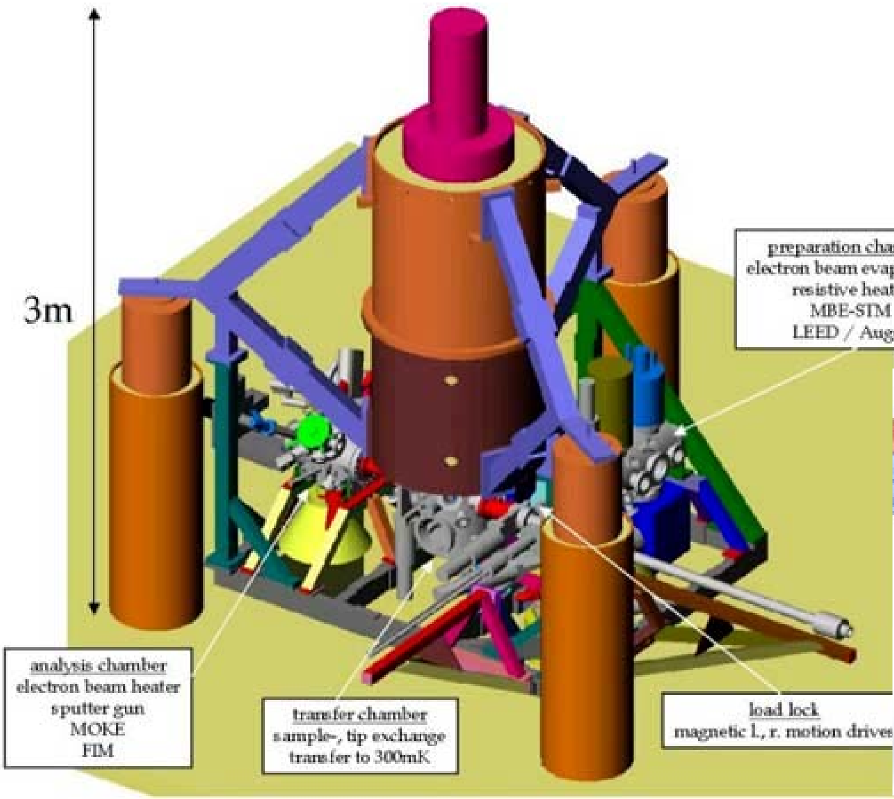






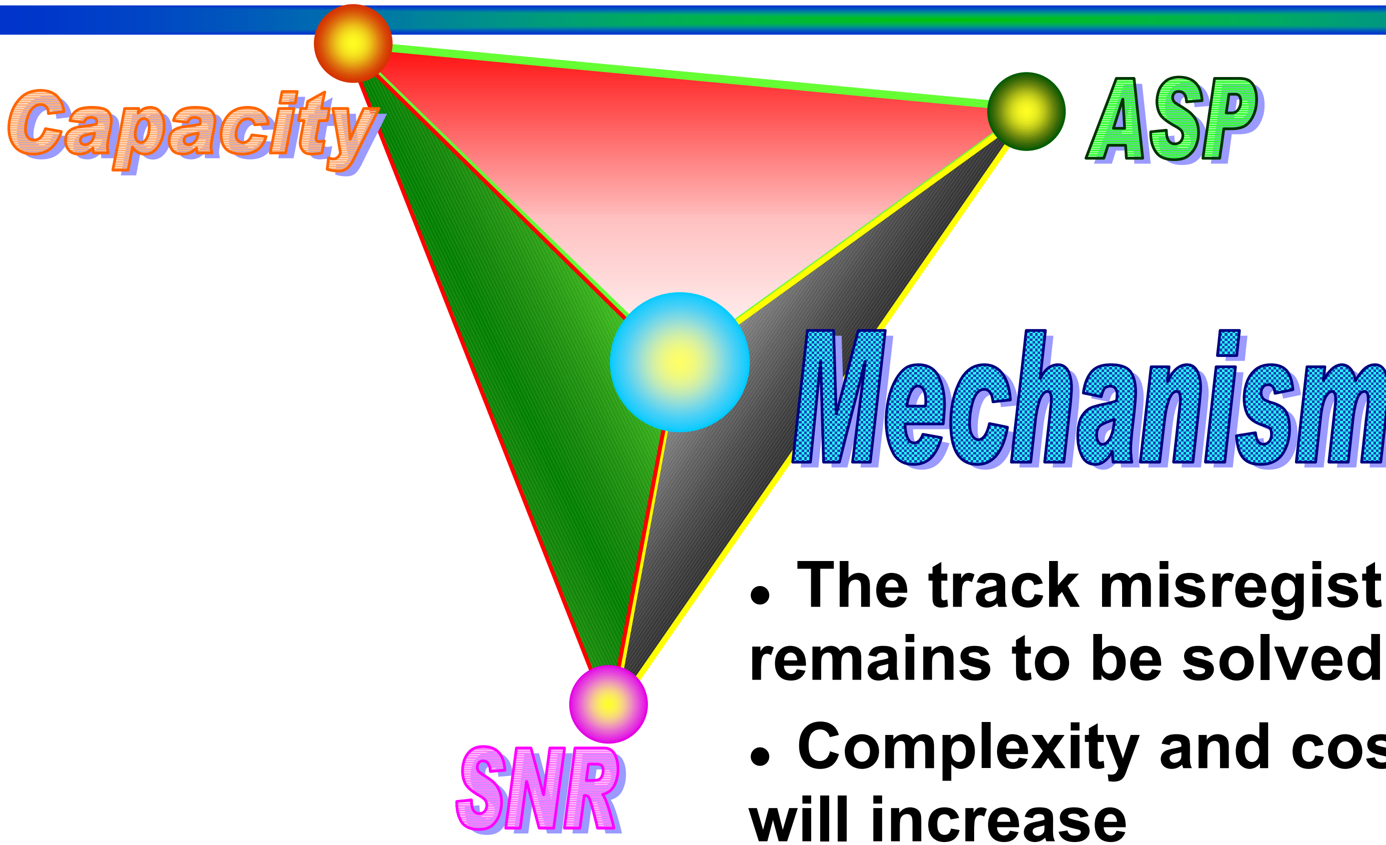

Two-atom-thick islands of Fe on W (110) with magnetization pointing either upward (light color) or downward (dark color.) The islands are a few nanometers across. A domain wall appears in the island left of center. Courtesy Prof. R. Wiesendanger, Hamburg

# … the drive is … “just engineering”




Ready for Dell PC’s?

# SNR, ASP, Capacity, Mechanism




- The track misregistration at 1 Tb/in$^2$ remains to be solved
- Complexity and cost of mechanism will increase
- Rotational vibration in enterprise market
- Head-media separation, how?

# Performance Issues at High AD

- Poor BER
- Low SNR
- Extreme Mechanical Tracking Requisites

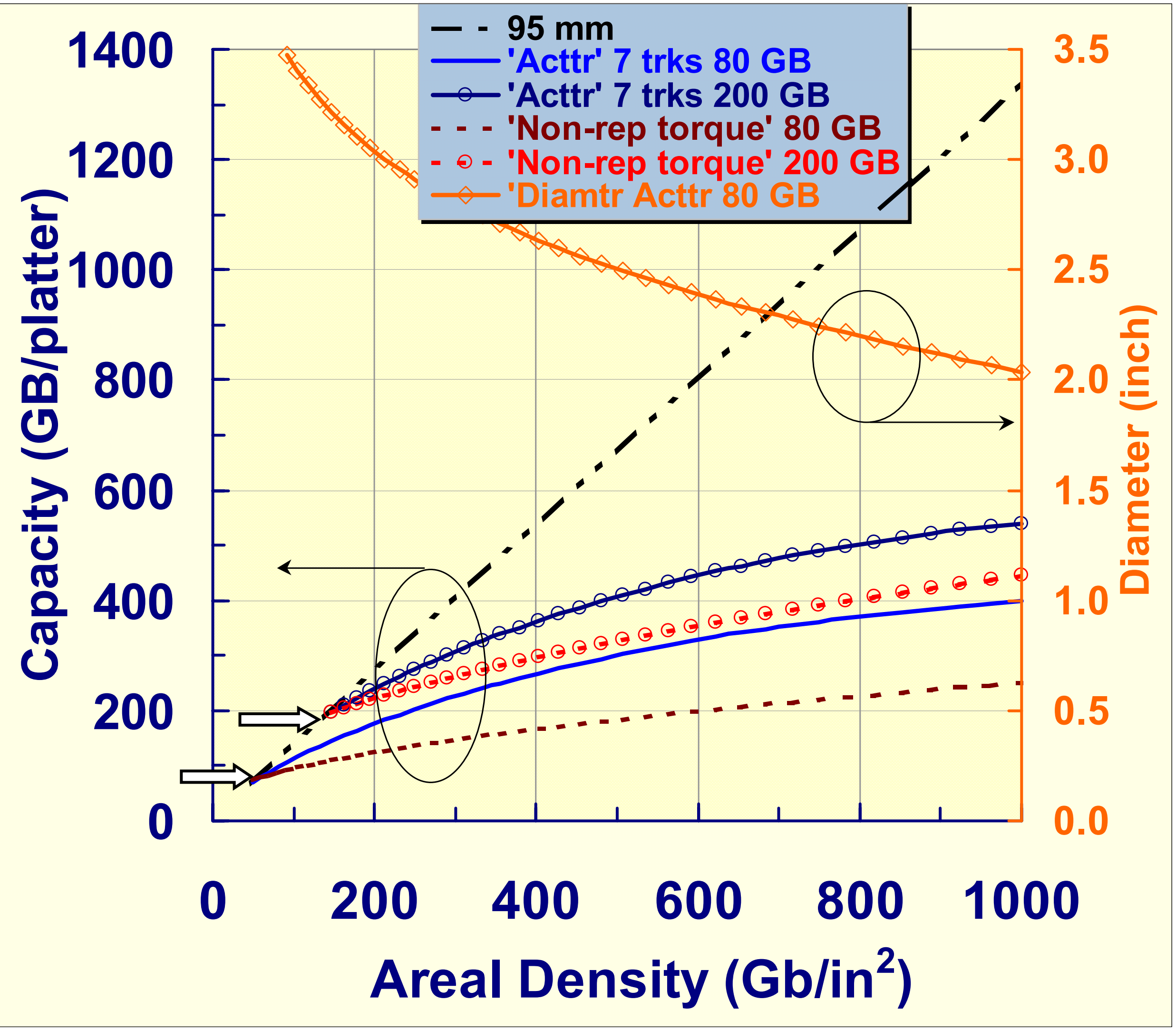
*These Factors Limit the Box Capacity Growth Even as Magnetic Areal Density Grows*

# Performance Issues at High AD

- Poor BER ← Increases ECC Overhead
- Low SNR ← Increases Servo Overhead
- Extreme Mechanical Tracking Requisites } ← Smaller Mechanical Devices

*These Factors Limit the Box Capacity Growth Even as Magnetic Areal Density Grows*

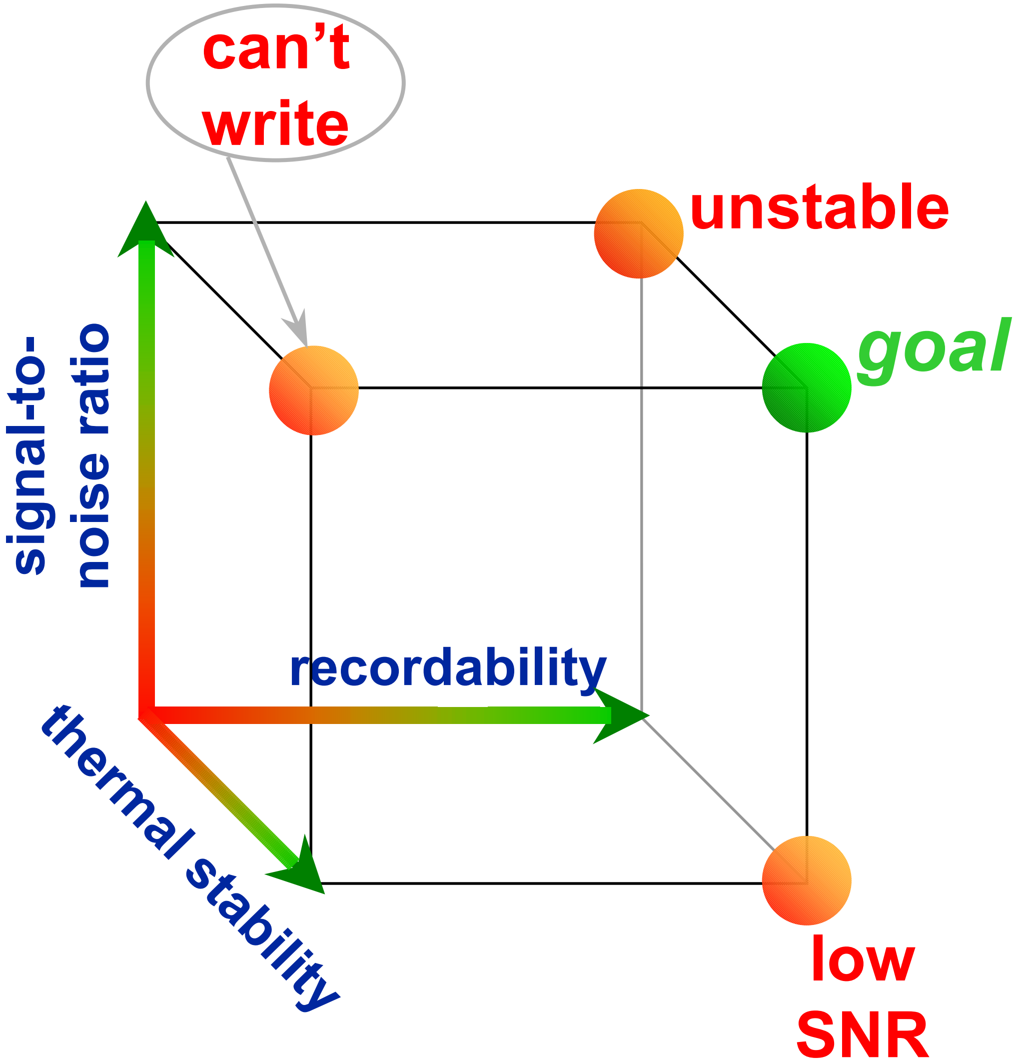
# Capacity vs. Areal Density




- BAR: 7.2 → 4
- $\tau$ : 10 → 5 ms
- 1 Tb/in² may not be cost-effective for producing a higher capacity drive with today's low price and ruggedness

See: G. Tarnopolsky, Trans. Magn. 40, No. 1, pp. 301-306 (2004)

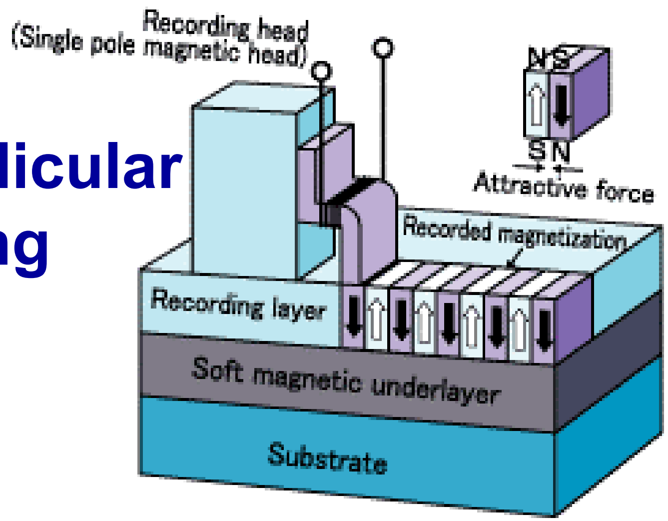
# Magnetic Areal Density Tradeoffs

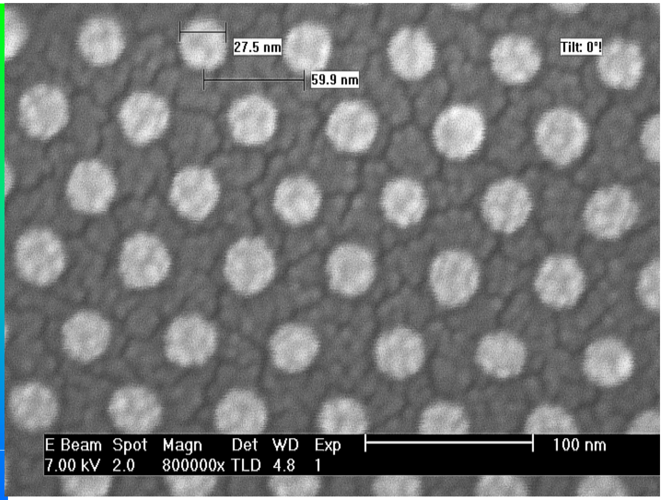



- INSIC's Extreme High Density Recording program strives for concurrent high SNR, permanency of the recorded bit, and ability to record (EHDR)
- INSIC's Heat Assisted Magnetic Recording program uses heat to achieve recordability (HAMR)
- Patterned media
- Tilted perpendicular media
- Self-organized media

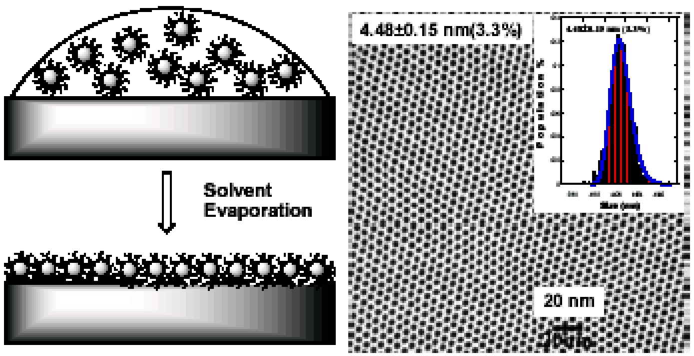
# HDD Technologies for the future

## Perpendicular recording



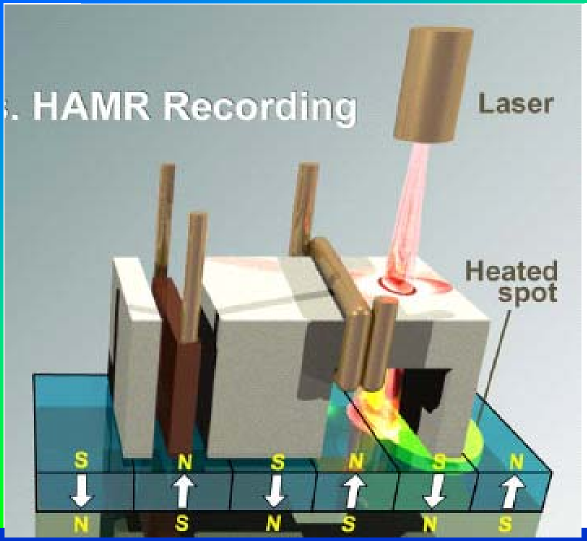




**Patterned media**, 33 nm bits. Bruce Terris, HGST




## Self-organized magnetic arrays

D. Weller, Seagate




**Heat-assisted magnetic recording** uses both laser and field to record T. McDaniel, Seagate

# Disk Drives at the Boundaries circa 2004

# Disk Drives at the Boundaries: Hi End

| Maker | Model | Format | Capacity | platters/heads | Areal density, max | Interface | Interface Datarate | Sustained transfer rate | Spindle speed | Latency | Seek/ average | Power operational/idle | operating shock | Sector size |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | mm | GB | | Gb/in2 | | Mbyte /s | Mbyte /s | rpm | ms | ms | watts | G | bytes |
| Hitachi Global | Desk-star 7K400 | 95 | 400 | 5/10 | 61.7 | ATA100/ SATA | 100/ 150 | 30/ 61 | 7,2k | 4.17 | 8.5 | 30/9.6 | 55 | 512 |
| Hitachi Global | Ultra-star 10K300 | 95 | 300 | 5/10 | 61 | Ultra SCSI & FC | 320/ 200 | 47/ 89 | 10k | 2.99 | 4.7 | 32.9/ 11.2 | 15 | 512 |
| Fujitsu | MAT300 | 95 | 300 | 4/8 | 75 | Ultra SCSI | 320/ 200 | | 10k | 2.99 | 4.5 | - /9.6 | 65 | 512 |
| Seagate | Cheetah 15K.3 | 65 | 73 | 4/8 | 34.7 | Ultra SCSI & FC | 320/ 200 | 49/ 75 | 15k | 2 | 3.6/3.9 | 16/12 | 60 | 512 |
| Seagate | Savvio | 65 | 73 | 2/4 | 60E | Ultra SCSI & FC | 320/ 200 | 41/ 63 | 10k | 3 | 4.1/4.5 | - /8.0 | 60 | 512 |

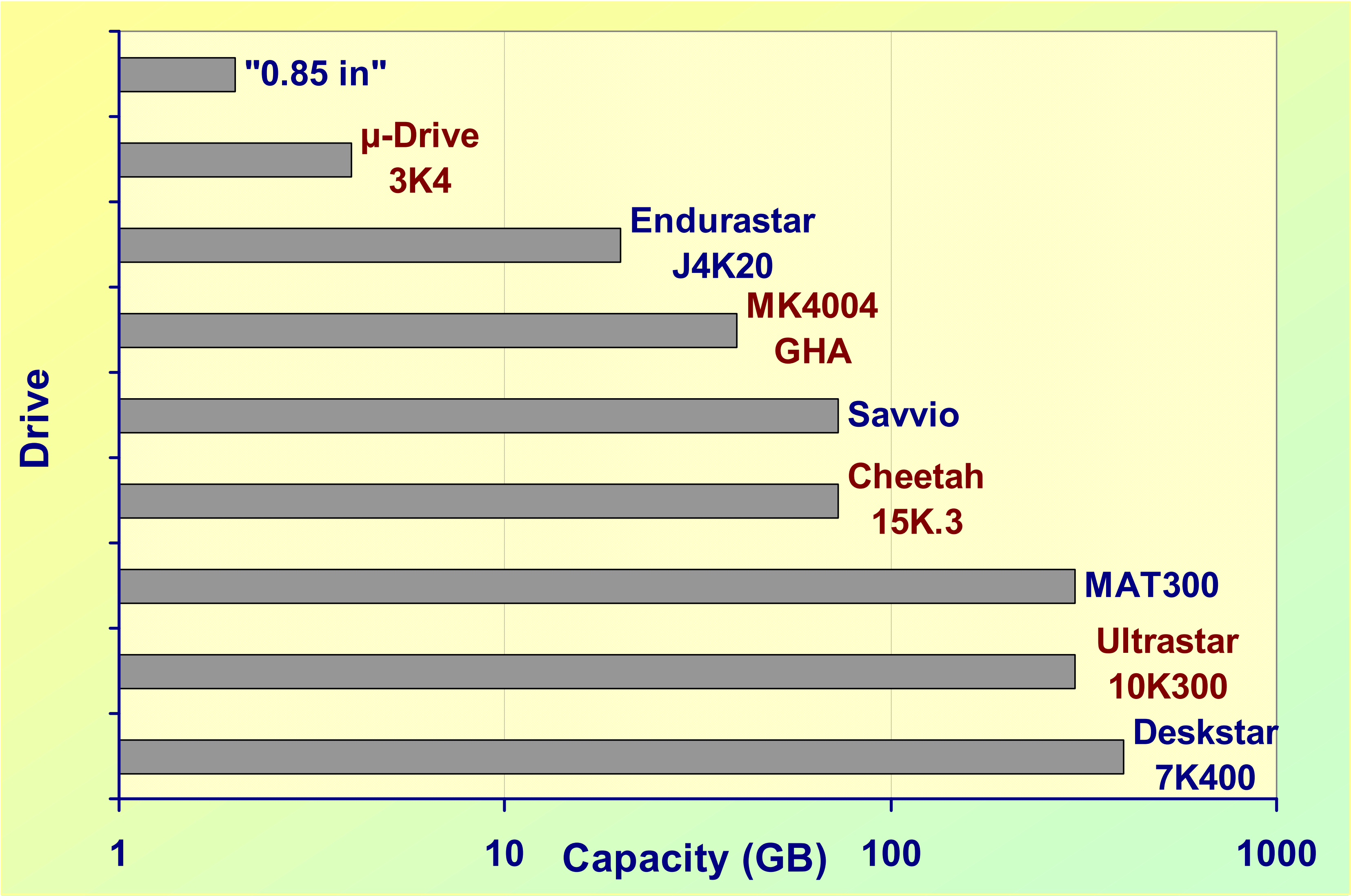
**Safe harbor: Representative examples, not a comprehensive industry compilation**

# Disk Drives at the Boundaries: Specialty

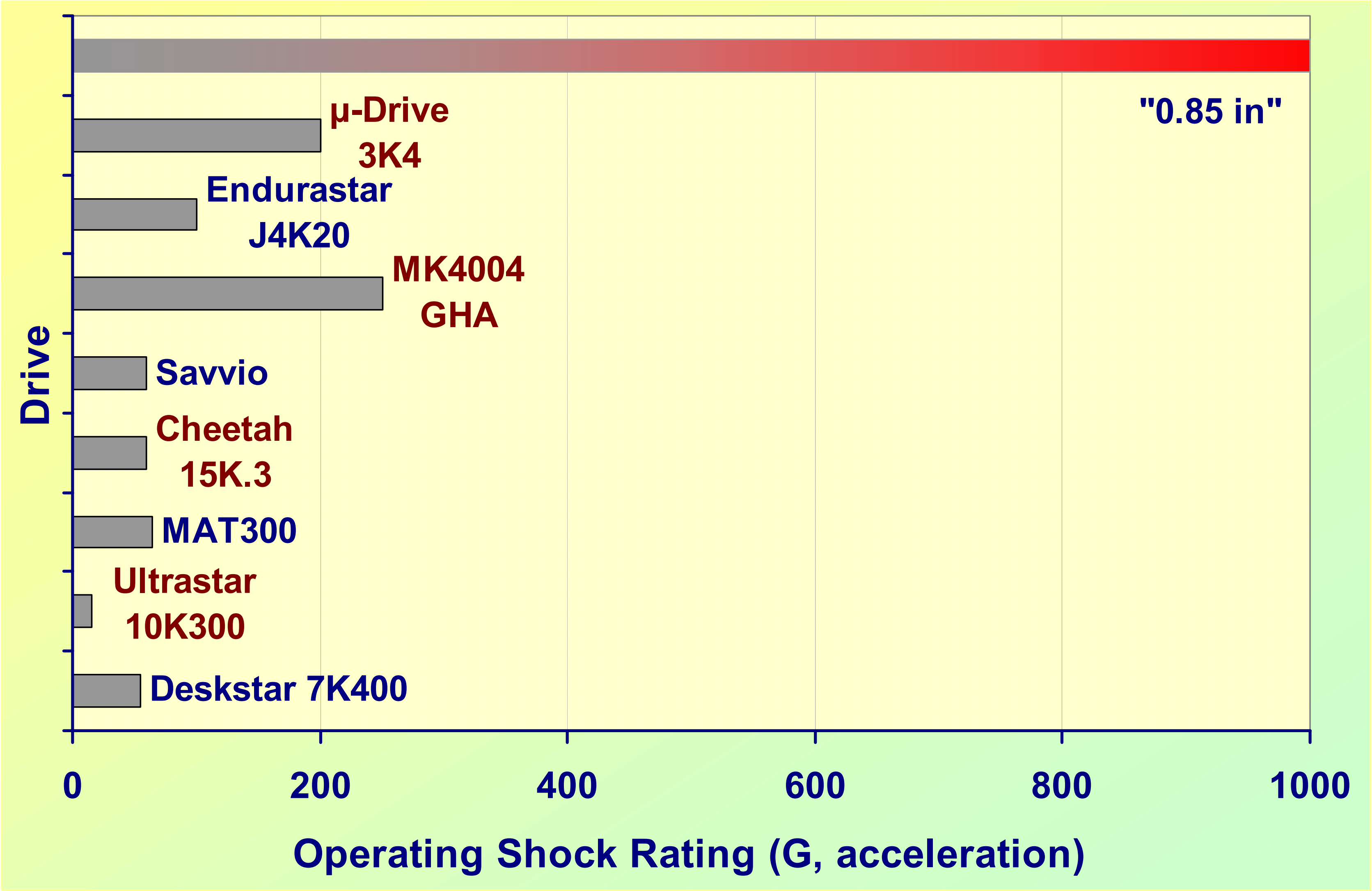
| Maker | Model | Format | Capacity | platters/heads | Areal density, max | Interface | Interface Datarate | Spindle speed | Latency | Seek/ average | Power operational/idle | operating shock | Sector size |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | mm | GB | | Gb/in2 | | Mbyte /s | rpm | ms | ms | watts | G | bytes |
| Toshiba | MK4004 GHA | 48 | 40 | 2/4 | 61.2 | ATA 100 | 100 | 4,200 | 7.14 | 15 | 1.4/ 0.08 | 250 | 512 |
| Hitachi Global | Endura star J4K20 | 65 | 20 | 1/2 | 37.7 | ATA 100 | 100 | 4,172 | 7.2 | 13 | 2.4/1.8 | 100 | 512 |
| Hitachi Global | µ-Drive 3K4 | 25.4 | 4 | 1/2 | 56.5 | ATA 33 | 33 | 3,600 | 8.3 | 13 | 1.2/ 0.23 | 200 | 512 |
| Toshiba | "0.85 in" | 22 | 2 | 1/1 | 60E | ad hoc | | 3,600 | 8.3 | | | 1000 | |

**Safe harbor: Representative examples, not a comprehensive industry compilation**

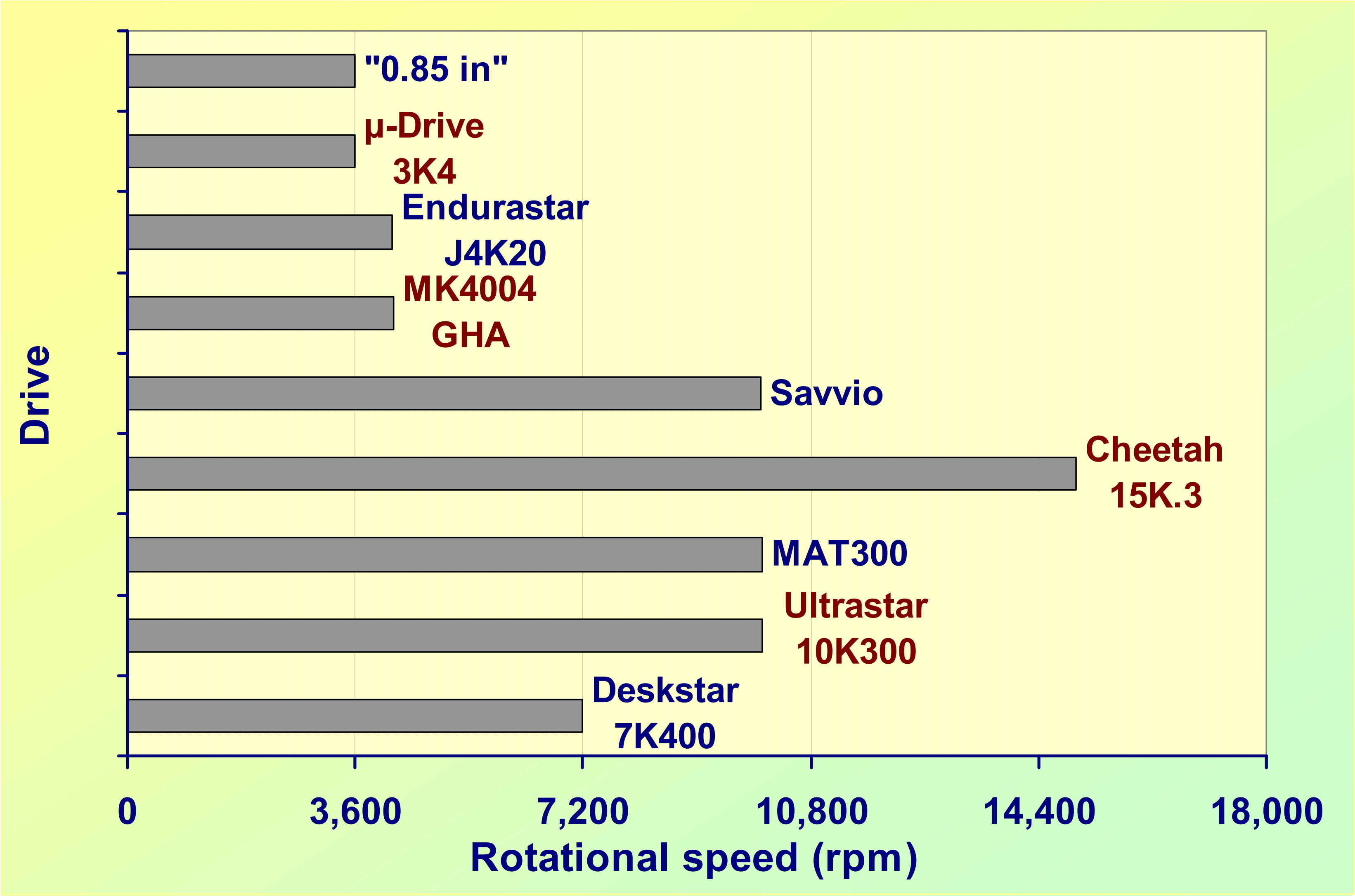
# Disk Drives at the Boundaries



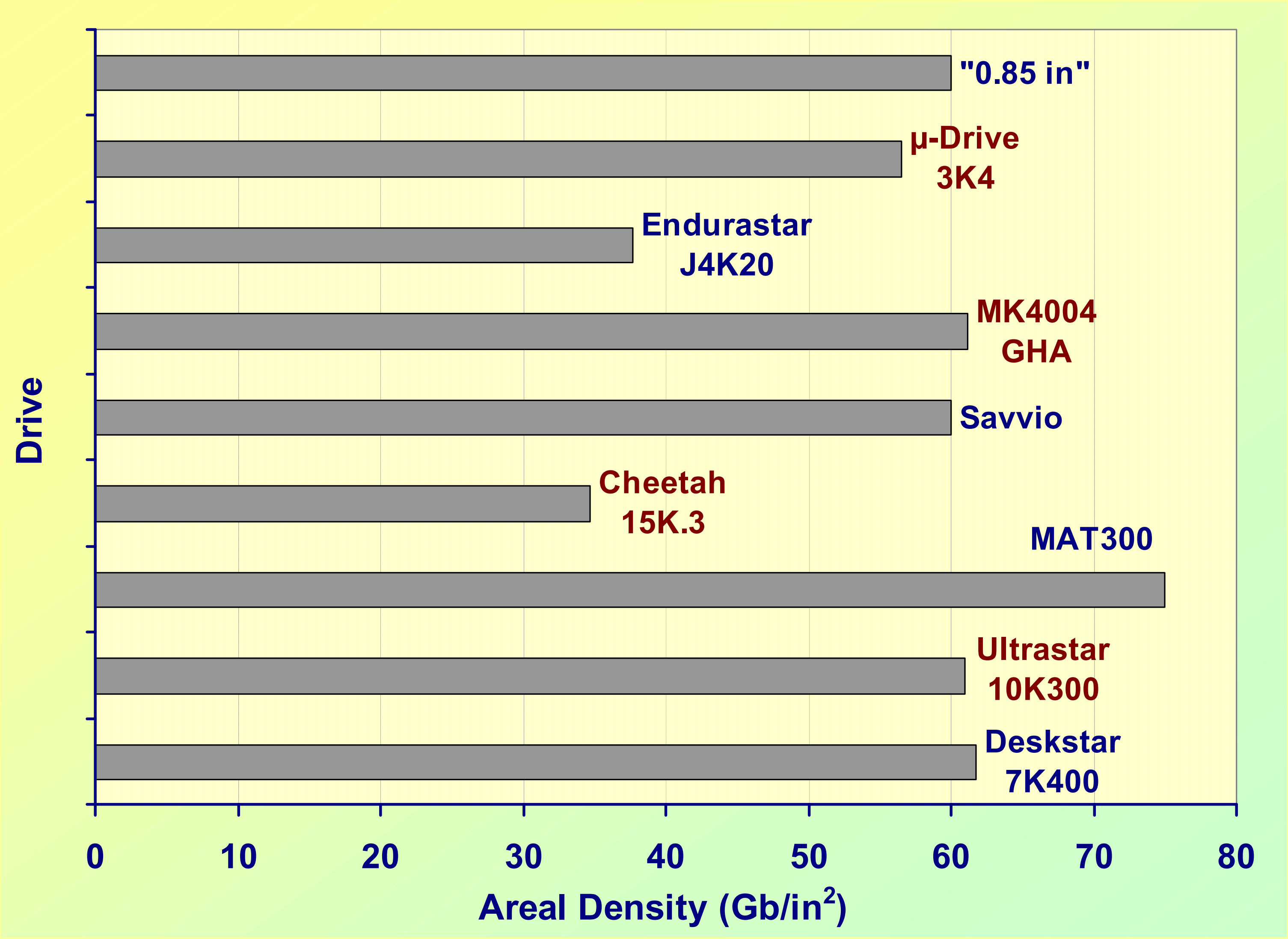
# Disk Drives at the Boundaries: Shock



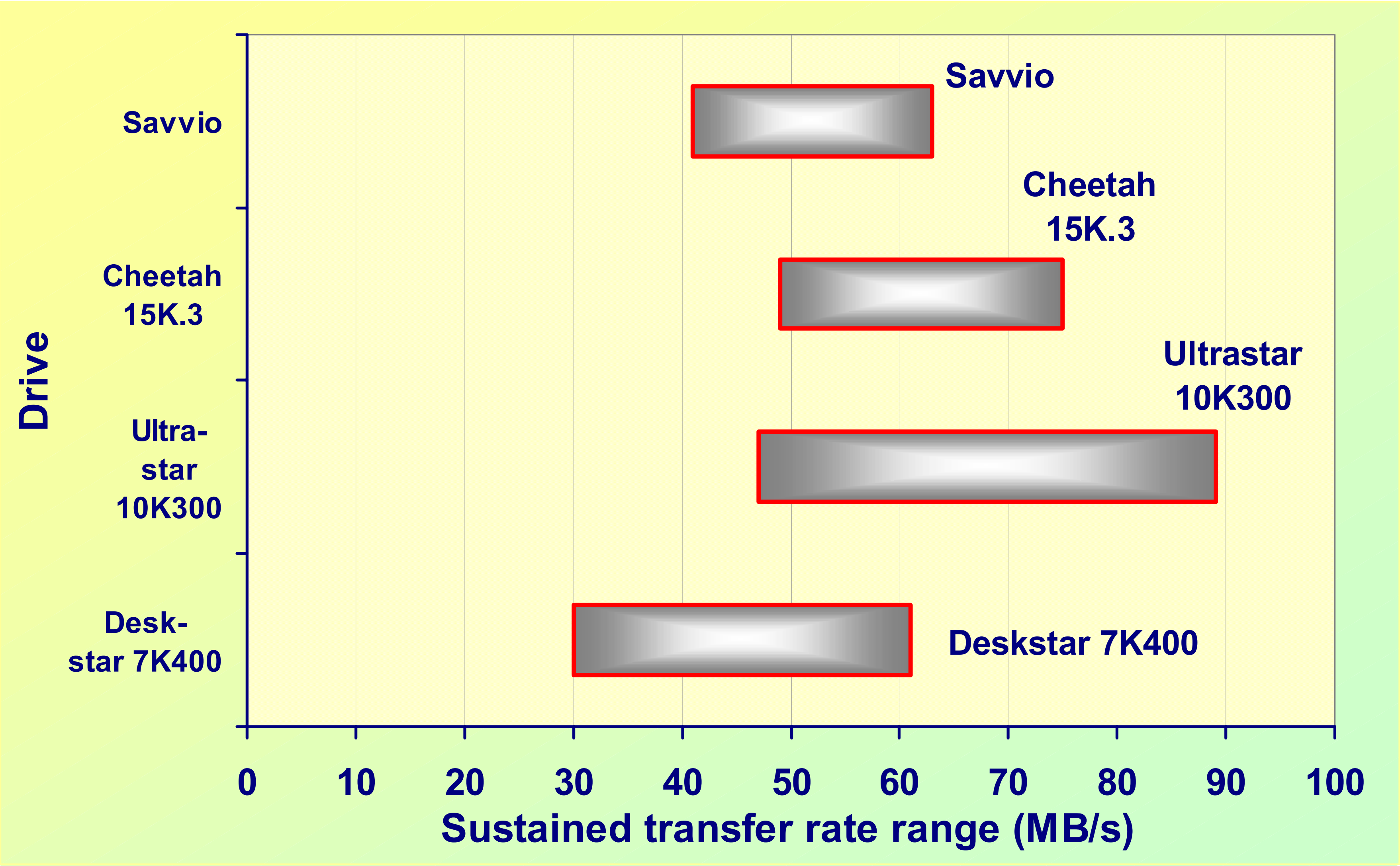
# Disk Drives at the Boundaries: RPM's



# Disk Drives at the Boundaries: AD



# Disk Drives at the Boundaries: MB/s



# Parting shots in "SAN Essentials"

- *Storage Area Networks Essentials*, Richard Barker and Paul Massiglia, John Wiley & Sons, pgs. 379-380, © 2002

> - "Since the evolution of SCSI in the mid-1980s, the functional definition of a disk *(or a tape)* has been essentially constant … the basic model of a single vector of numbered blocks has remained the same."
> - "Researchers today are questioning whether the tried-and-true disk *(or tape)* functional model is the most effective way to provide network storage services."
>
> *(text in italics added by GT)*

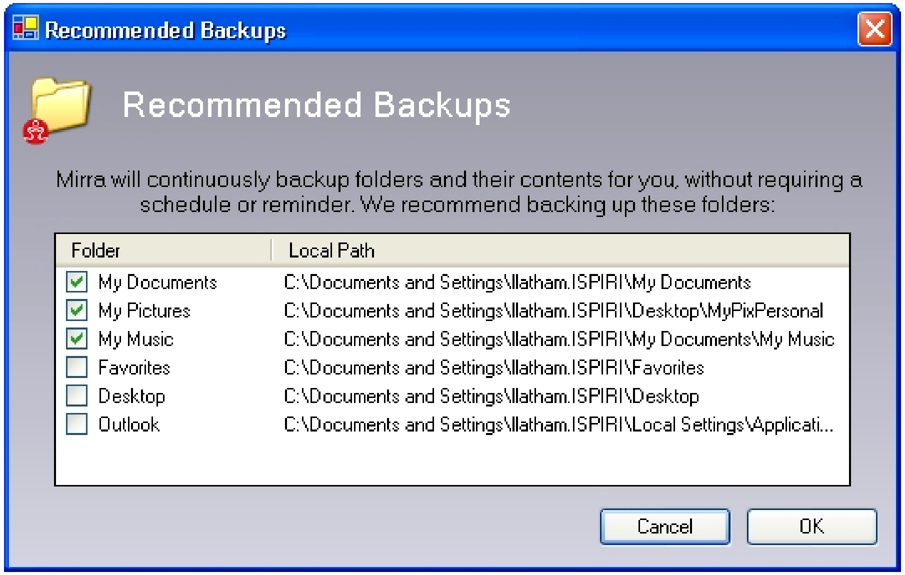
# Personal NAS at 80 to 160 GB

- **MIRRA Server**
- 80 GB/$400 or 120 GB/$500
- Ethernet/USB, file sharing
- Secure
- Auto backup

- The value of data management is relevant for *all* markets, not just the enterprise market

- 80 GB/$640, 120 GB/$870, 160 GB/$920
- Ethernet
- File sharing

- SnapAppliance's Snap Server 1100




***What about capacity?***

- Prices as of 12/6/2003

# Market Evolution

## Limited but very high capacity per spindle

≥***300 GB*** drives being offered in the consumer markets
(300 GB Maxtor MaxLine II 5,400 rpm, Ultra ATA 133, $ 250,
***400 GB*** HGST 7K400),

and the enterprise markets
(300 GB Fujitsu 10,000 rpm, Ultra 320 SCSI & 2 Gb/s FC,
300 GB HGST 10,025 rpm , Ultra 320 SCSI or 2 Gb/s FC)

## Proliferation of specialty systems

"Tiny" drives, iPod drives, rugged drives, personal NAS, ...

# INSIC & Data Storage Devices and Systems Research

## Information Storage Industry Consortium

# Information Storage Industry Consortium

- collaborative research consortium for the worldwide information storage industry
- established 1991

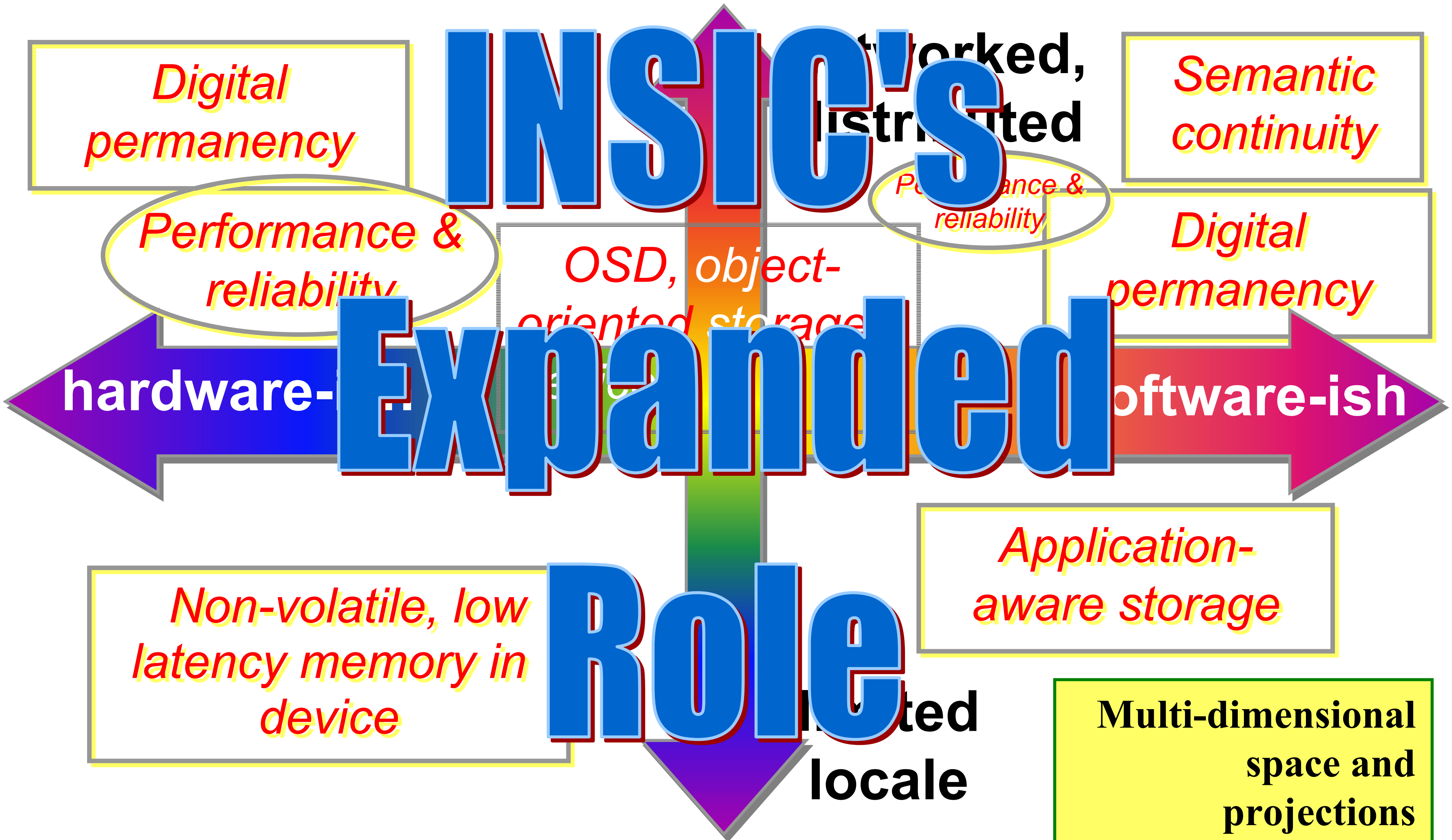
- Conduct Joint Research on High Risk Pre-competitive Storage Technologies
- Develop Technology Roadmaps
- Maximize Value of University Research
- Obtain Government Funding
- Speak for the Industry

IBM
ECD*
IDC*
SONY
MAXELL
IMATION
APRILIS*
QUANTUM
SAMSUNG
CERTANCE
MAGNECOMP*
STORAGETEK
DOWA MINING*
MEMS OPTICAL
AGERE SYSTEMS
WESTERN DIGITAL
TORAY INDUSTRIES
HEWLETT PACKARD
VEECO INSTRUMENTS
ADVANCED RESEARCH
SEAGATE TECHNOLOGY
EUXINE TECHNOLOGIES
HUTCHINSON TECHNOLOGY
HITACHI GLOBAL STORAGE TECHNOLOGIES

* Limited Member

FLORIDA INTERNATIONAL UNIVERSITY
DATA STORAGE INSTITUTE (DSI)
LOS ALAMOS NATIONAL LAB
CENTRAL LANCASHIRE
COLORADO STATE
JOHNS HOPKINS
NORTHEASTERN
UC SAN DIEGO
MANCHESTER
UC BERKELEY
OHIO STATE
COLORADO
PLYMOUTH
MISSOURI
NEBRASKA
VIRGINIA
ALABAMA
HARVARD
ALBERTA
ARIZONA
ILLINOIS
MIT
ISIC
NIST
IDEMA
IDAHO
MITRE
PURDUE
STANFORD
MINNESOTA
VANDERBILT
SANTA CLARA
GEORGIA TECH
ARIZONA STATE
NORTHWESTERN
CARNEGIE MELLON
ARGONNE NAT'L LAB
WASHINGTON UNIVERSITY
LAWRENCE BERKELEY NAT'L LAB
LAWRENCE LIVERMORE NAT'L LAB
NATIONAL UNIVERSITY OF SINGAPORE

# Storage Devices and Systems Research



# DS2 Workshop, UCSD, April 27-29

- DS2 = *Data Storage Devices and Systems*
- Purpose:
  - Establish a technology roadmap in the comprehensive space of systems and devices
  - Answer the question whether there are pre-competitive research topics in data storage systems, where industrial cooperation and joint sponsorship of academic research is not preempted by market competition. This favors research into difficult, high-risk, or long-term issues

# DS2 Brainstorming Thrusts

- Application-aware storage
- Active storage devices
- Privacy and security
- Autonomic storage
- Long-term storage
- Pervasive storage

***Technical Committee***

Paul Frank INSIC
Craig Harmer Veritas
Paul Massiglia VERITAS
Paul Siegel CMRR/UCSD
Giora Tarnopolsky INSIC
James Hughes StorageTek
Michael Mesnier Intel/CMU
Thomas Ruwart DTC U Minn.
Erik Riedel Seagate Research
Gordon Hughes CMRR/UCSD
Remzi Arpaci-Dusseau U Wis.

# Archiving Problem is Growing

- Government regulation and business necessity
- Archive growth outpaces storage growth
  - More data, longer retention
  - Expanding scope
    - e-mail, instant messages, trader conversations
- Current solutions do not address
  - Regulatory compliance
  - Data organization and search
  - Changing data formats
  - Media obsolescence
  - Multiple media types
  - Security issues

**Fred van den Bosch / Veritas**

# Long-term Storage: ≥ 10 years

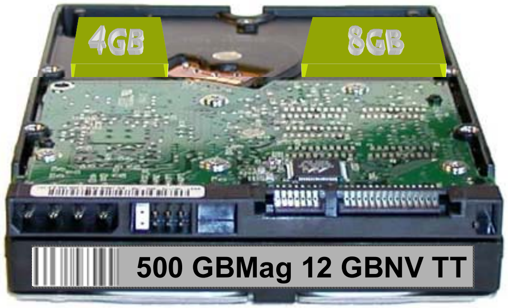
- There is no assured scheme for perpetual content preservation
  - Semantic continuity: make computer languages evolve like natural languages, assure comprehension
  - Storage management independent of the medium and of the content itself

- Systems hold "eternal" data in devices bearing a ~ three-year warranty[§,*]
- Digital assets have undergone migrations to devices of higher performance & volumetric density. *No more.*

Safe harbor: §) H/W MTBF ≥ 30 yr.
*) Tape cartridges guaranteed "for life"

# Large, non-volatile storage stratum

- Multi-GB, ≤ µs-latency, non-volatile memory
- Flash, M-RAM, or MEMS
- Non-volatile stratum would be "virtual disk" such as disk is "virtual tape" to tape system
- Various application archetypes could have assigned non-volatile storage streams - until objects of associated types are transferred to specific areas on the media
    - Intelligent space allocation, self-defragmentation



# OSD - Object-based Storage Device

- OSDs take the *storage-device-specific* component of the file system into the storage device itself
- Ability of device to manage its own capacity
- Ability of device to export file-like objects to their clients
- Where in the storage hierarchy is the OSD concept to be applied? A 400 GB device used to be a RAID. Now it is a single drive.

# Concluding Remarks: Endurance of Good Ideas

# Permanency of good ideas

- Antikythera Mechanism
- An astronomical mechanical artifact recovered from a vessel that sunk circa 80 BCE by the Antikythera Island, Greece
- The mechanism contains a *differential gear*, an invention that was lost then, to be re-invented about 1,500 years later!

# Disk Drives - Perfect Invention

- 2-D travel with only one linear motion
- High volumetric density
- Random access
- Mass-produced
- Non-volatile
- Affordable
- Rugged

Few-hundred $/box
No vibration isolation
no T stabilization
These properties define drives

**Broad spectrum of new and enhanced applications**

# Functionality at High Capacity

- **The "Four R's" of HDD's:**
  - Reliability — Terabytes last for ever (mag & mech)
  - Ruggedness — Field devices, not lab curios
  - Remote — All data remote: Mobile & networked
  - Readability — Find the file in the haystack
- **Make data-handling function match the raw capacity: enhanced OSD, self-managing, ...**
- **Vast opportunities for useful devices at ~200 $Gb/in^2$, high SNR, low ECC overhead**
- **Proliferation of applications**