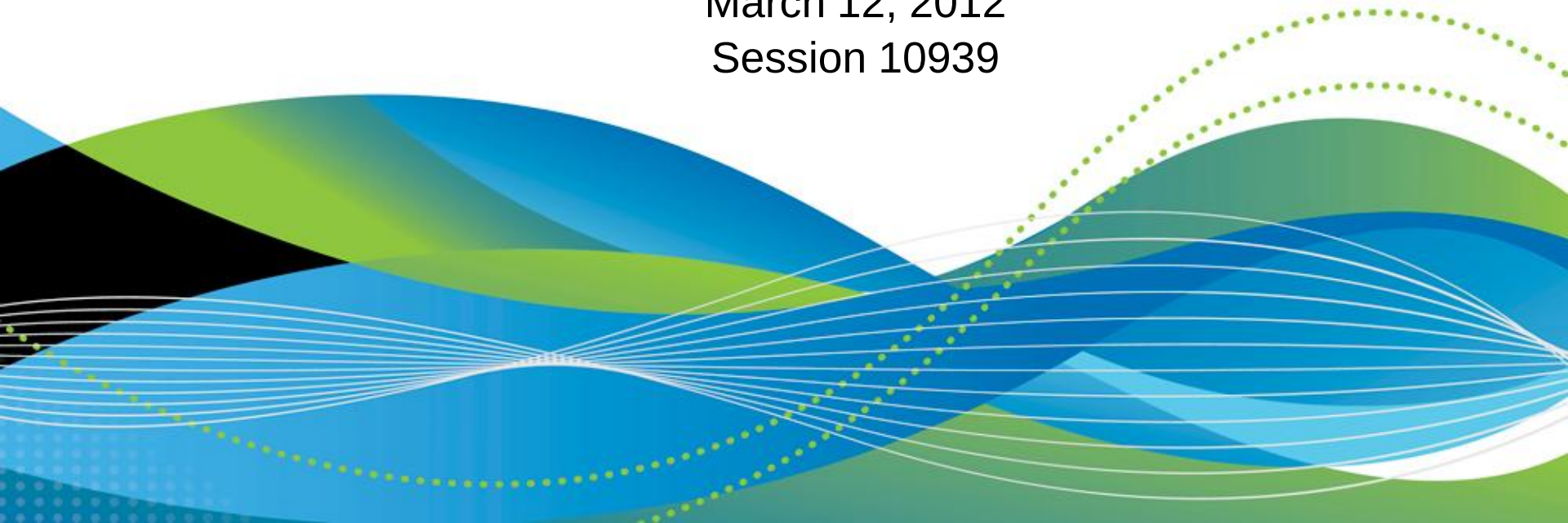
#SHAREorg

SHARE
Technology · Connections · Results

# Storage: The Insatiable Demand….when does it end?

Gary Francis
Oracle

March 12, 2012
Session 10939



# Agenda

- Storage Demand Drivers
- Cost / Areal Density Trends
- A Few Comparisons
- Technology Opportunities
- Summary

PRESENTATIONS

FINANCIAL REPORTS

BACKUPS

MP3s

SOFTWARE CODE

PHOTOS

REGULATORY COPIES

# ~2 ZETTABYTES in 2011

DATA MINING

VIDEOS

ARCHIVE

TRADEMARKS

DATABASES

EMAIL

EMPLOYEE RECORDS

LEGAL DOCUMENTS

R&D COPIES

CAD DRAWINGS

Source: Storage Magazine '11

# 2 Zettabytes Equals….

- Every person in the US @ 3 tweets per minute, non-stop, for 26,976 years
- Every person in the world having 215 million high-resolution MRIs daily
- 200 billion 2-hour HD movies which would take one person 47 million years (24x7) to watch
- Enough information to fill 28 billion 64GB iPads…which could:
  - Build the Great iPad Wall of China – 2x the average height
  - Cover 86% of Mexico City
  - Build a mountain 25x higher than Mt. Fuji (12,000 feet tall)

Source: Horison Information Strategies

# Data Storage Growth

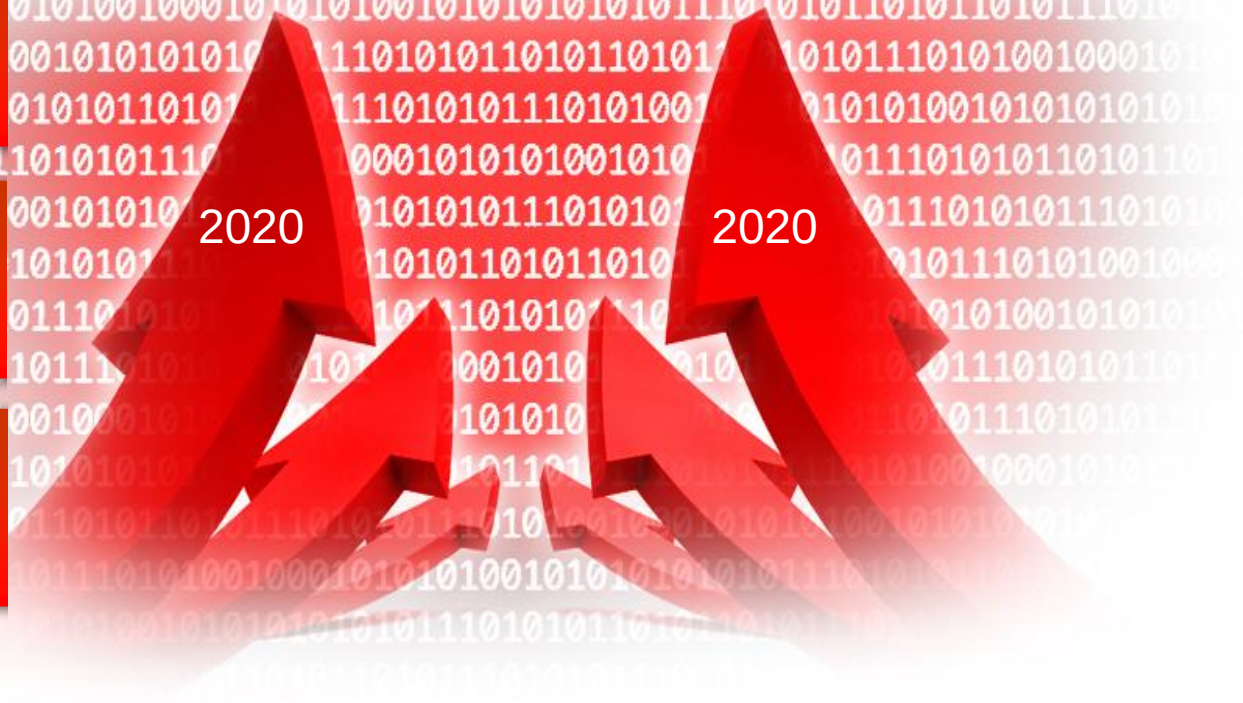
The Nature of Storage is Changing

35,000 Exabytes

Digital Data is Expected to Grow to 35,000 Exabytes by 2020

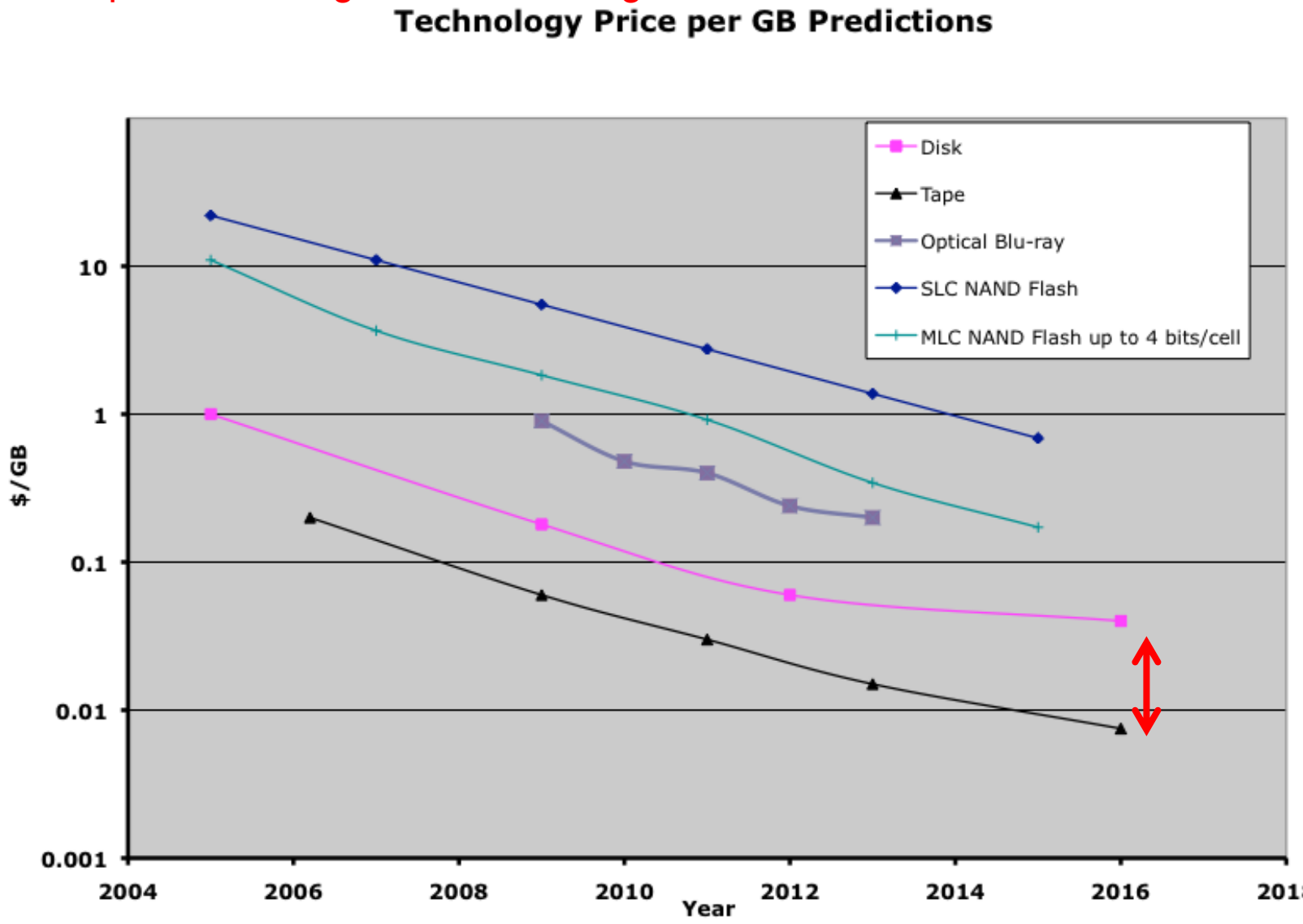
Storage Management Costs More

Storage Consumes ~40% of Data Center Power



# Technology Price/GB Projections

…and Tape's Advantage Is Accelerating



# It's Not Only About Cost/TB

| | Disk | Tape |
|---|---|---|
| Max shelf life (bit rot) | 10 years | 30 years |
| Best practices for data migration to new technology | 3-5 years | 8-12 years |
| Uncorrected Bit Error Rate, Probability (avg 1 error in x TB) | $10^{-14}$ ( ~10's of TB) | $10^{-19}$ (~1 million TB) |
| Power and cooling | 238X | X |

Each technology refresh or migration has a cost associated with it

# Long Term Cost Advantage

"Tape is always much less expensive than disk and always uses much less power when measured on a per petabyte basis."

- Disk is 15 times more expensive than tape
- Disk uses 238 times more energy for an archiving application with 45% annual growth over a 12 year period
- Disk energy consumption alone costs more than ALL the costs of tape!

From "In Search of the Long-Term Archiving Solution — Tape Delivers Significant TCO Advantage over Disk", The Clipper Group, December 2010

# 60 Years of Storage Innovation

Megabytes to Zettabytes


## In the beginning

- Tape – 1951 / Univac UNISERVO
  - 1.4MB / 7.2KB/sec
  - Metal tape – nickel plated phosphors bronze
  - Protein robots

- Disk – 1956 / IBM RAMAC
  - 5MB – 100 bpi – 1200 RPM
  - 50 24” disks
  - 5’ long, 5’8” tall, 29” deep
  - Weight > One Ton

# 60 Years of Storage Innovation

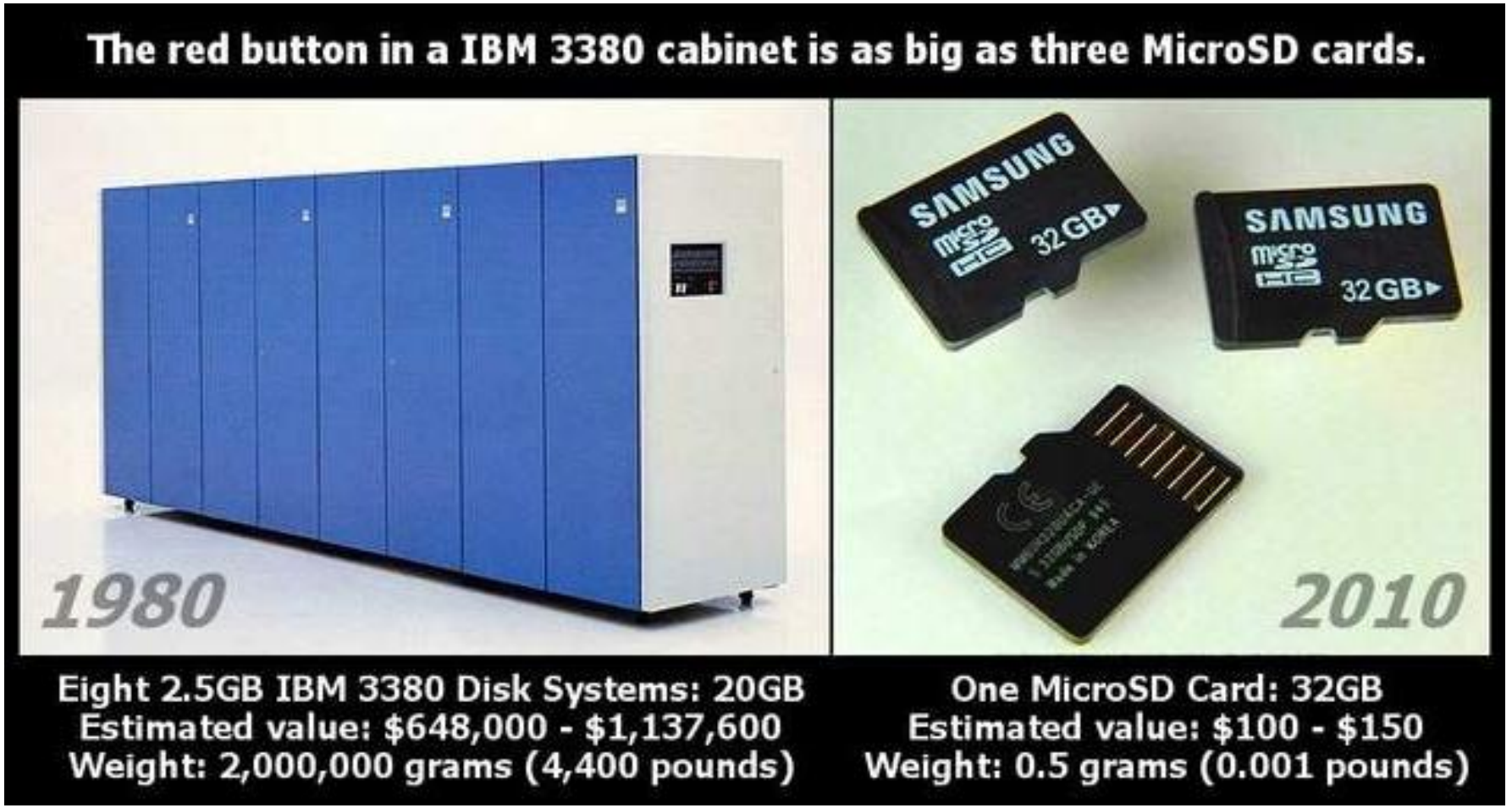
Megabytes to Zettabytes

## And then there was real storage

- Tape
  - The capacity leader – Oracle 10000C
  - 5TB uncompressed
  - 240 MB/sec
  - Robotics have advanced somewhat ☺


- Disk
  - The areal density leader: Toshiba MK7559GSXP
  - 500GB per 2½” platter
  - 744 Gb/in²

# A Disk Comparison



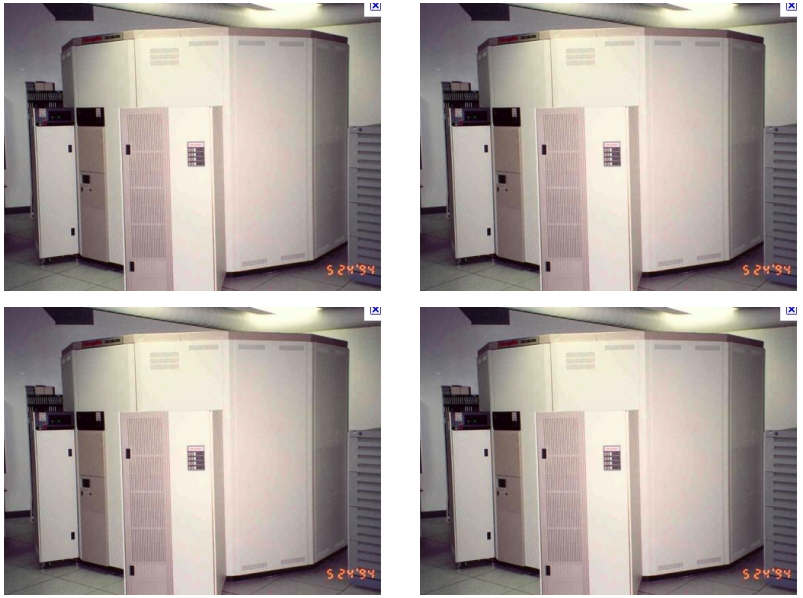
# A Tape Comparison

## Oracle StorageTek T10000C


- 5TB capacity uncompressed
- 252MB/s data rate
- 2 drives = 480MB/s

=

## StorageTek 4410 Automated Cartridge System



- 5,000 4480 carts
- 5TB capacity uncompressed
- 80 drives = 360MB/s data rate

# And with 2:1 Compression


=

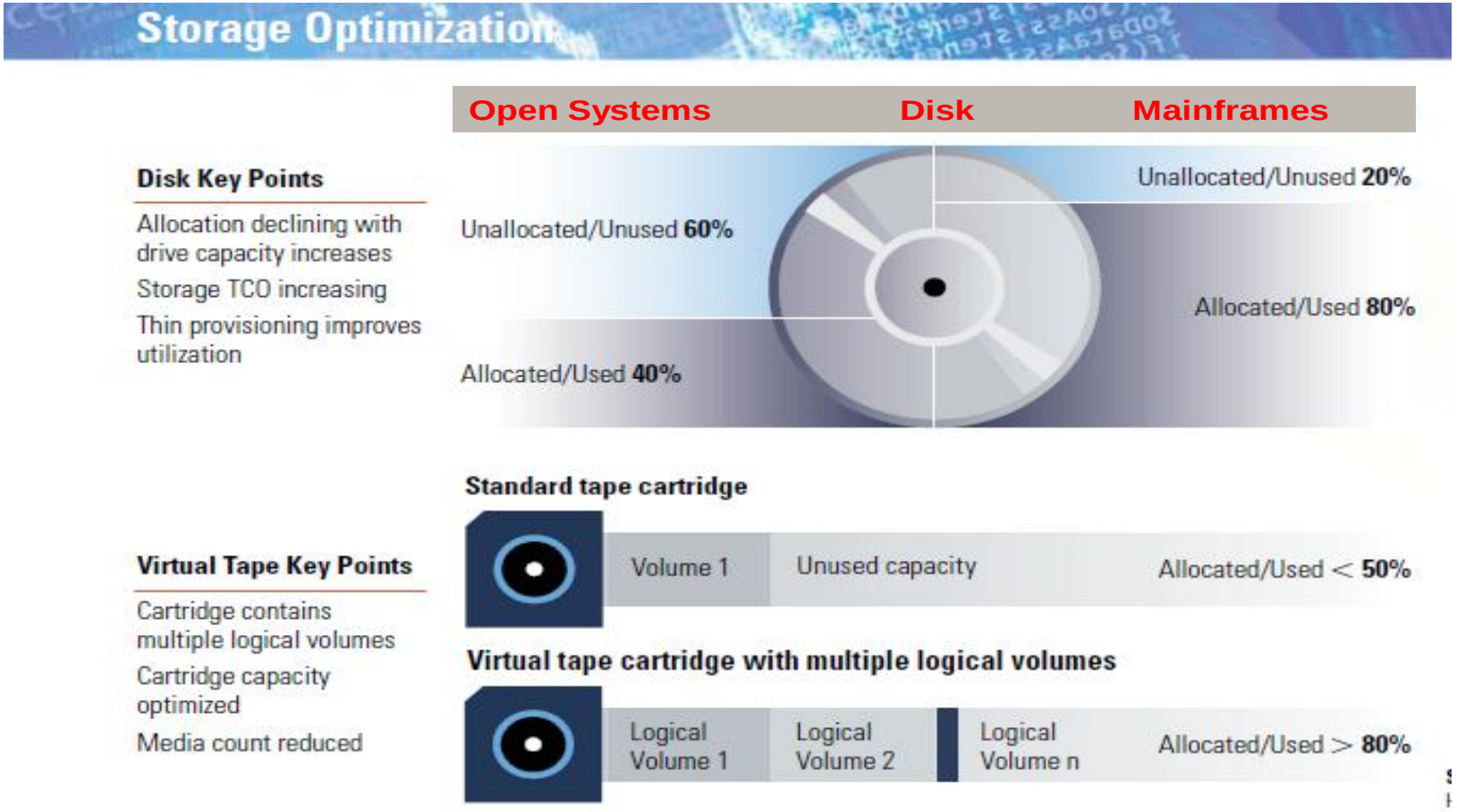






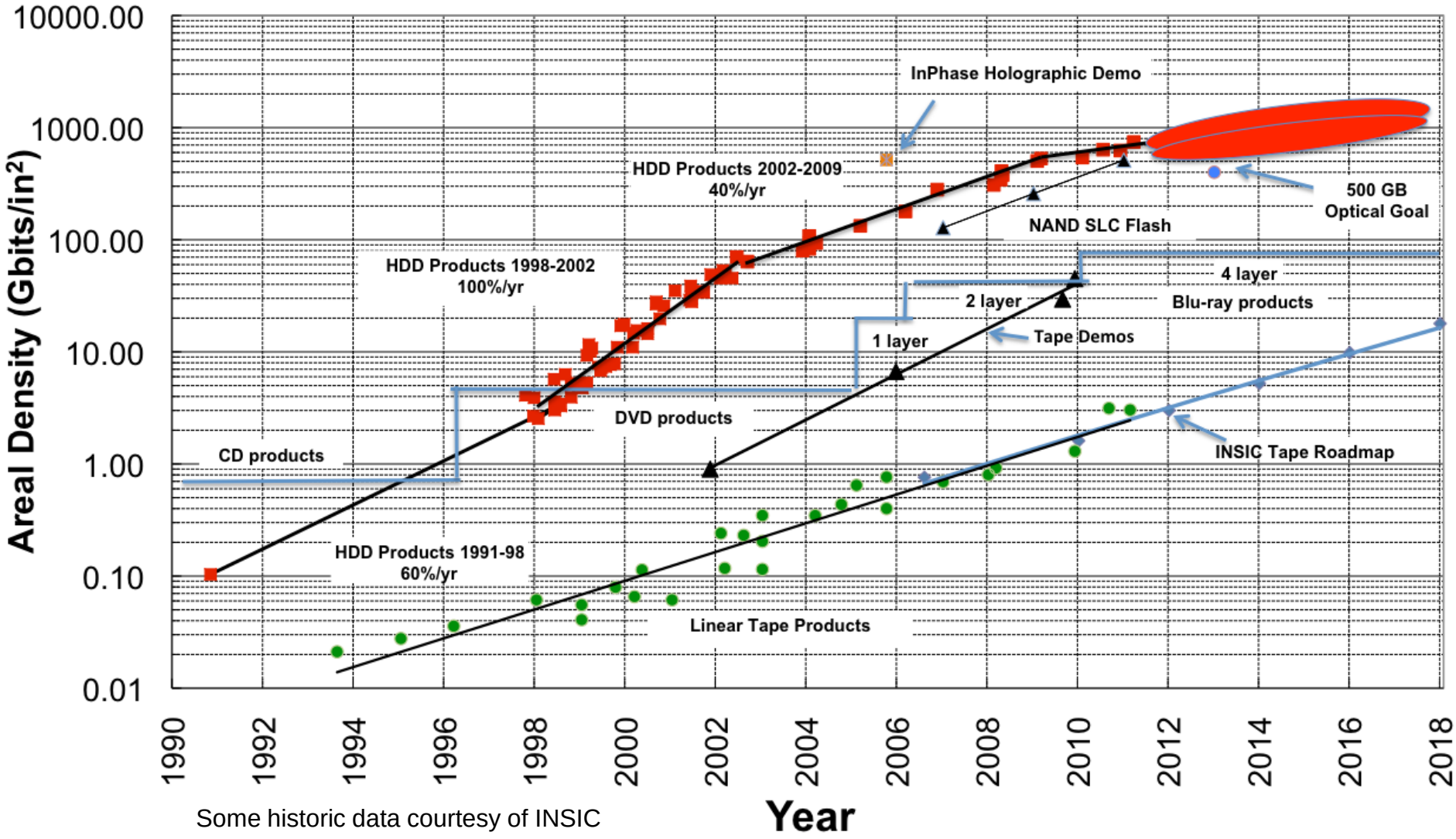



# Mainframe Storage




Source: Horison Information Strategies

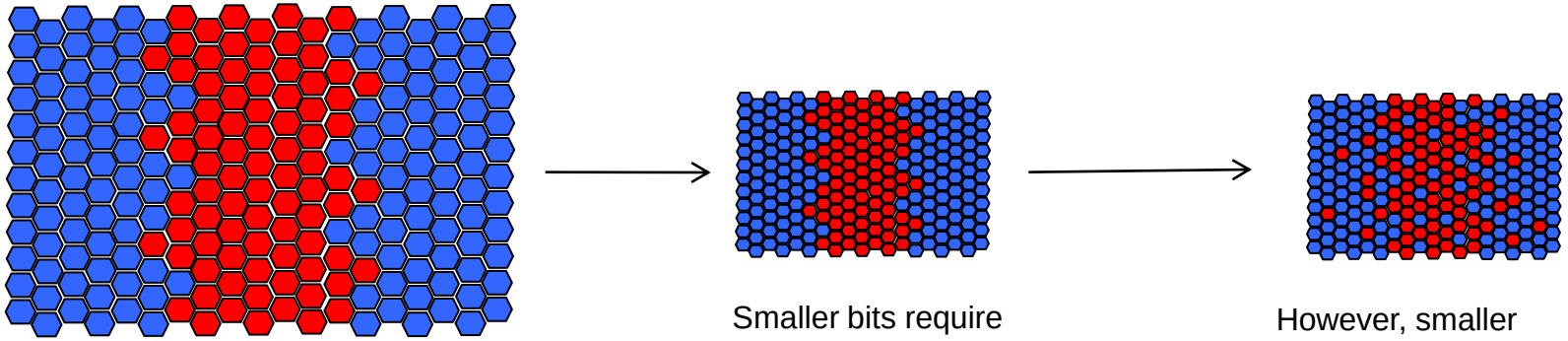
# Storage Technologies Areal Density Trends




Some historic data courtesy of INSIC

Tape gets its capacity by having 1000X the recording surface area comparing a 1/2 inch cartridge to a 3 1/2 inch disk.

# Super-Paramagnetic Effect



Smaller bits require smaller grains for required SNR

However, smaller grains have a higher probability of reversing over time

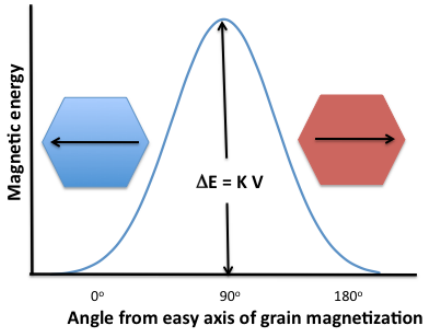
Neel-Arrhenius law gives: Mean time to randomly flip grain due to thermal fluctuations

$$\tau_N = \tau_0 \exp\left(\frac{KV}{k_B T}\right)$$

V is the volume of the grain, T is the temperature and K is the grain's magnetic anisotropy energy

$\frac{KV}{k_B T}$ > 60 for good thermal stability, 10 year data life [1,2]

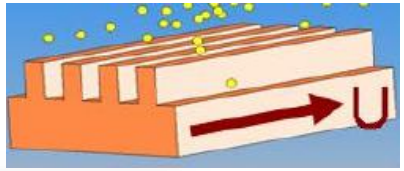
$\frac{KV}{k_B T}$ > 90 for today's tape [3], 30 year data life

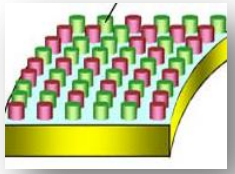




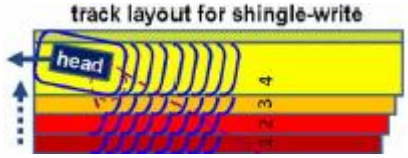
1. Dobisz et al. Patterned Media: Nanofabrication Challenges of Future Disk Drives, Proceedings of the IEEE, Vol. 96, No. 11, November 2008
2. Weller et al. Thermal Effect Limits in Ultrahigh-Density Magnetic Recording, IEEE Transactions on Magnetics, VOL. 35, NO. 6,November1999
3. Watson et al. Investigation of Thermal Demagnetization Effects in Data Recorded on Advanced Barium Ferrite Recording Media, IEEE Transactions on Magnetics, Vol. 44, No. 11, November. 2008

# Possible New Disk Technologies





- Discrete Track Recording (DTR) (~2 Tb/inch$^2$)
  - Pre-defined tracks = higher TPI
- Bit Patterned Media (BPMR) (~5 Tb/inch$^2$)
  - Pre-defined single grain domain bits
- Heat Assisted Magnetic Recording (HAMR) (~5 Tb/inch$^2$)
  - Laser heats media to assist grain reversal
- Microwave Assisted Magnetic Recording (MAMR) (~5 Tb/inch$^2$)
  - Oscillating magnetic field in head causes media magnetic domains to switch at lower head magnetic fields
- Shingled Recording (SWR) (~5 Tb/inch$^2$)
  - Overlapping /overwriting of tracks




- 2-D Recording (TDMR)
  - Improved coding and signal processing algorithms (two dimensional)
- HAMR + BPMR -> 10 Tb/inch$^2$

Source: “Future Options For HDD Storage” Y. Shiroishi, et al, IEEE Trans. On Mag. Vol. 45, NO 10, Oct 2009

# Disk Observations

- HDD constraints likely until 4Q12
- High-capacity HDDs are GREAT!!!!.....but :
  - Performance is not improving
  - Don't forget about rebuild times
  - Access Density problems need to be solved vs. under utilization
- New consortium within IDEMA (ASTC advanced storage technology consortium) of disk companies and suppliers started 1/31/2011 to address development of new disk technologies
  - New technologies not ready yet
  - Focus on capacity and not performance
- More and more consolidation…from 60 companies in 1980
  - Western Digital to buy Hitachi HDD announced 3/7/2011 – completed 3/8/2012
  - Seagate to buy Samsung HDD announced 4/19/2011 - completed 12/19/2011
  - Seagate 40% HDD shipment share[4], WD (50%) and Toshiba (10%)


1. http://www.techpowerup.com/149967/Toshiba-Boosts-Performance-with-New-High-Areal-Density-1-TB-2.5-Inch-Hard-Drive.html
2. http://www.theinquirer.net/inquirer/news/2047313/seagate-unveils-3tb-external-hard-drive
3. Barry C. Stipe, et al. "Thermally Assisted Magnetic Recording on Granular & Patterned Media at up to 1 Tb/in2" INSIC 2010 annual meeting Aug 4.
4. Source IDC

# Tape Storage Projections

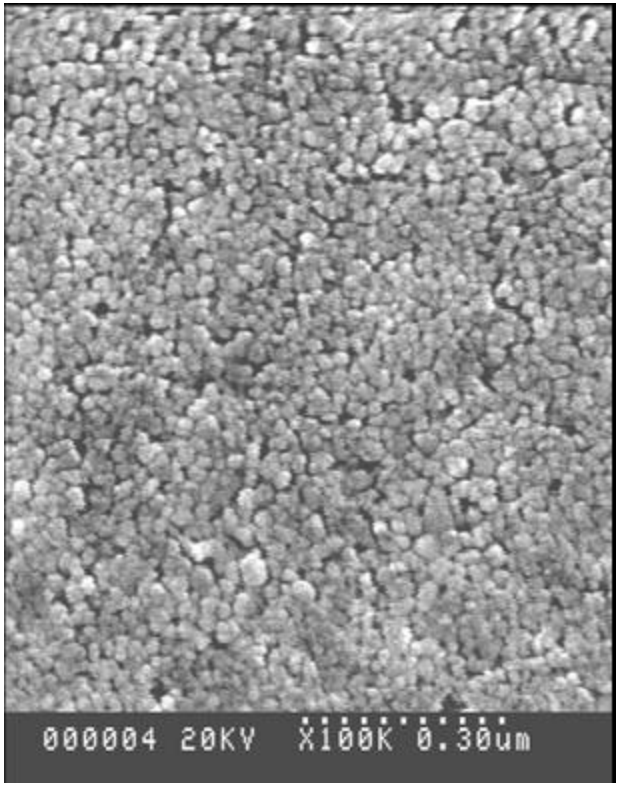
- Expect cartridge capacities to double at approximately 2 year intervals (T10KC breaks this trend with 5X and TS1140 with 4X increase)
  - $/GB reduces by roughly ½ every two years
- Oracle and IBM engineering demos show technology for 10 TB cartridge capacities (1,2)
- INSIC tape roadmap shows technology path to 32 TB on a cartridge (3)
  - IBM and FujiFilm demo areal density to support up to 35 TB cartridge (4)
  - Hitachi Maxell, Ltd. and Tokyo Institute of Technology demo areal densities to support 50TB cartridge(5).
- Scaling could get up to 64 TB on a cartridge (6)
- Challenge is to increase track density with interchangeable flexible media


1. D. Berman, R. Biskeborn, N. Bui, E. Childers, R. D.Cideciyan, E. Eleftheriou, D. Hellman, et al., ‘‘6.7 Gb/in2 Recording Areal Density on Barium Ferrite Tape,’’ IEEETrans. Magn. 43, No. 8, 3502–3508 (2007).
2. Internal Sun documents
3. International Magnetic Tape Storage Roadmap, September 2008. Copies available on request.
4. http://www.zurich.ibm.com/news/10/storage.html
5. http://www.computerworld.com/s/article/9177396/Researchers_develop_nanotech_for_50TB_tape_cartridge
6. IBM Journal of Research and Development, Storage Technologies and Systems Volume 52, Number 4/5 2008, “Scaling tape-recording densities to 100Gb/inch2.

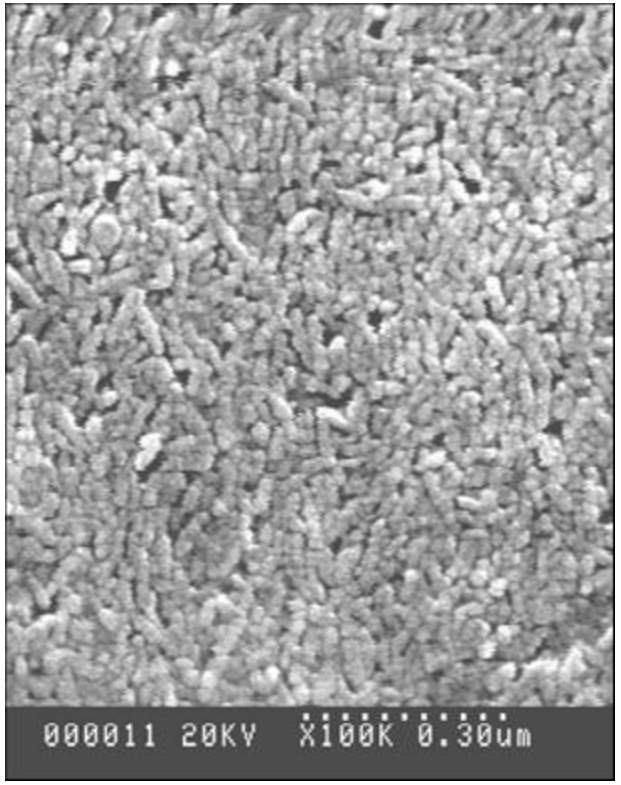
# Tape Technology Areas of Focus

- New Media (BaFe)
  - More stable than ever before
    - Magnetic (thermal), chemical, dimensional
  - Smaller particles or grains
  - Lower noise media structures
  - Bonus = reuse for future generations at rated capacity & performance
- New Recording Channels
  - New LDPC codes
- New Heads
  - MR (magneto resistive) >>>> GMR (giant magneto resistive)
  - 32 channels

# Magnetic Recording Surface



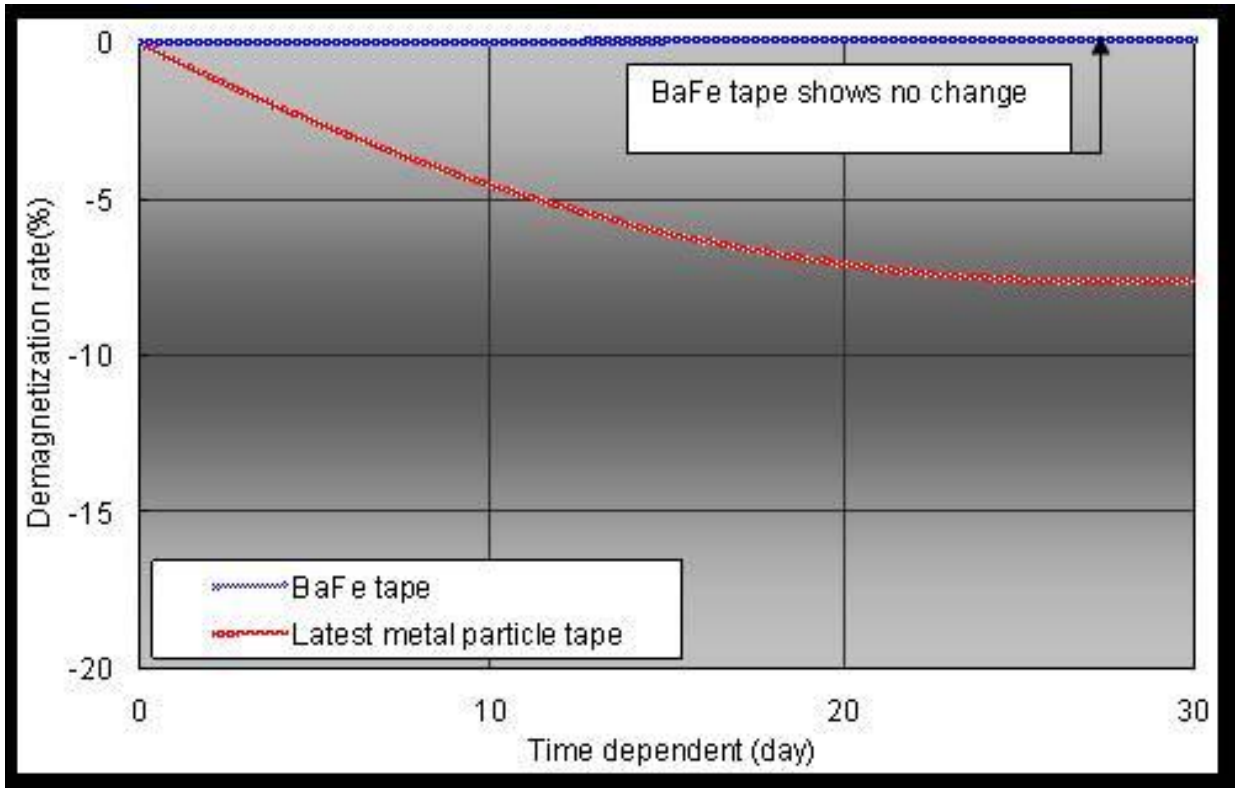

SEM of BaFe Media




SEM of MP Media

# Magnetic Degradation

## Measuring Changes in Demagnetization




- 60C/90% RH for 30 days (equivalent to 30 years ambient storage)
- No change in BaFe magnetic properties

http://www.fujifilm.com/news/n100910.html

# Tape File Systems (LTFS)

- LTFS stands for: Linear Tape File System[1,2]
  - Allows for a single tape cartridge to appear like disk
  - Files stored in hierarchical directory structure
  - File metadata is stored in a partition at the beginning of the tape cartridge
- LTFS format specification describes the layout of the data structures stored on tape
- LTFS Open Source client software writes and reads to the tape in the LTFS format
- Cartridges written in LTFS format are self describing
  - No special application is needed to see the files[3]


1. Pease, et al, "The Linear Tape File System," Proceedings of the 2010 IEEE 26th Symposium on Mass Storage Systems and Technologies
2. X. Zhang, et al, "HPTFS: A High Performance Tape File System," Proceedings of 14th NASA Goddard/23rd IEEE conference on Mass Storage System and Technologies, College Park, MD, May 2006.
3. Need Open Source LTFS Client code

# Flash Memory Observations

- Device types
  - SLC (single level cell) – Higher write cycles but higher $$/GB
    - 100K cycles
  - MLC (multi-level cell – 2 bit) – Lower $$/GB but lower write cycles
    - 1K – 3K cycles
  - TLC (multi-level cell – 3 bit) – Even lower $$/GB and even lower write cycles
    - 100 – 500 cycles
- 10x price differential over DRAM
- Advantages over magnetic disk
  - Access time / Transfer rate / Power consumption
  - Challenges
  - Limited write/erase cycles can be addressed by proper system architectures
  - How small can it go?  25nm to 19nm to 10nm

# World's Smallest Magnetic Storage

- Developed by scientists from IBM and the German Center for Free-Electron Laser Science
- Nanometre data storage built atom by atom
  - Aligned in rows of 6 atoms each
  - 2 atom rows = 1 bit
  - 8 pairs of atom rows = 1 byte
  - 4 x 16 nanometre
    - Nanometre = 1 millionth of a millimetre

- 100x greater than current HDD technology
- However....stability is reached at minus 268 degrees Centigrade

# Summary

- Disk areal density is slowing
- The gap between magnetic disk and flash is narrowing
  - Use for cache, tiered storage, high performance applications
- Tape is the ideal archive media and is widening the gap between disk
  - 15X less TCO than disk[1]
  - 238X less energy than disk[1]
- New tape architectures make using tape easier
  - Tape is more like disk with partitions and LTFS
  - Disk is more like tape with shingled recording formats
- New and innovative storage technologies will continue to be developed....feasibility TBD
- None of this will eliminate the need for improved storage efficiencies

1. “In Search of the Long-Term Archiving Solution —Tape Delivers Significant TCO Advantage over Disk”, Clipper Notes, December 23 2010, Report #TCG2010054LO. 1 PB archive growing at 45% per year over 12 years.