

Microsoft

**WinHEC**

2007

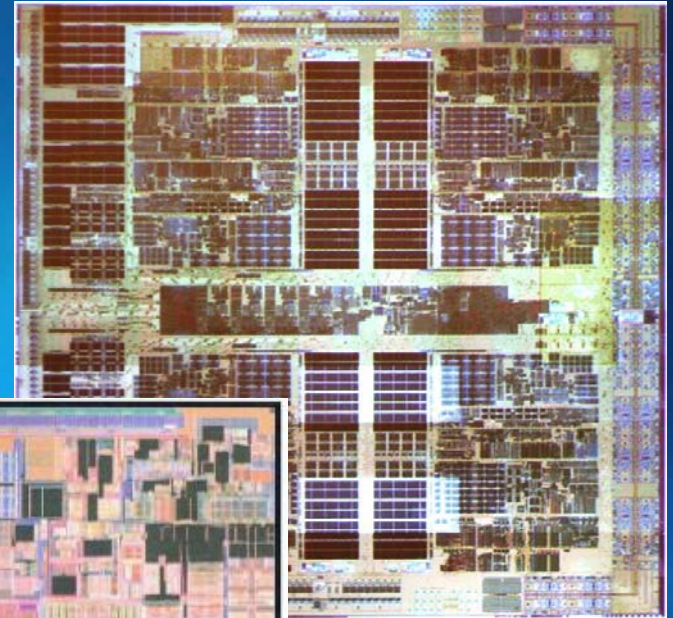
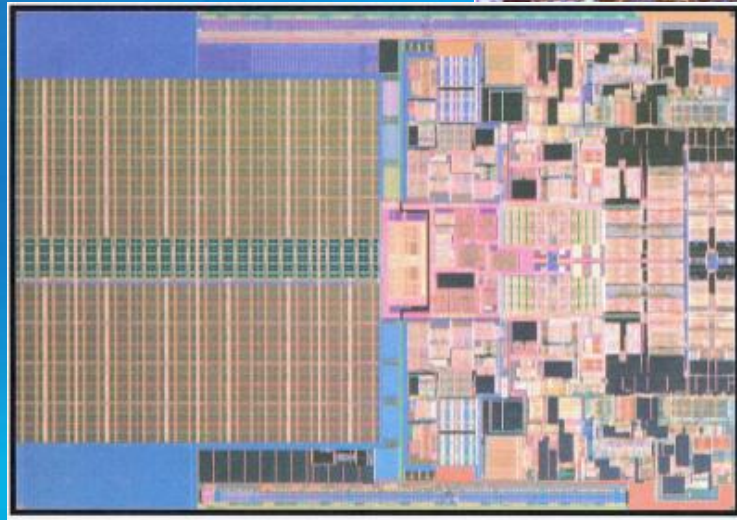


# The Future of Memory and Storage: **Closing the Gaps**

Dean A. Klein  
VP Market Development  
Micron Technology, Inc.

# Processor Trends

- Increasing core performance
- Increasing cores
- Increasing bus speed
- More memory



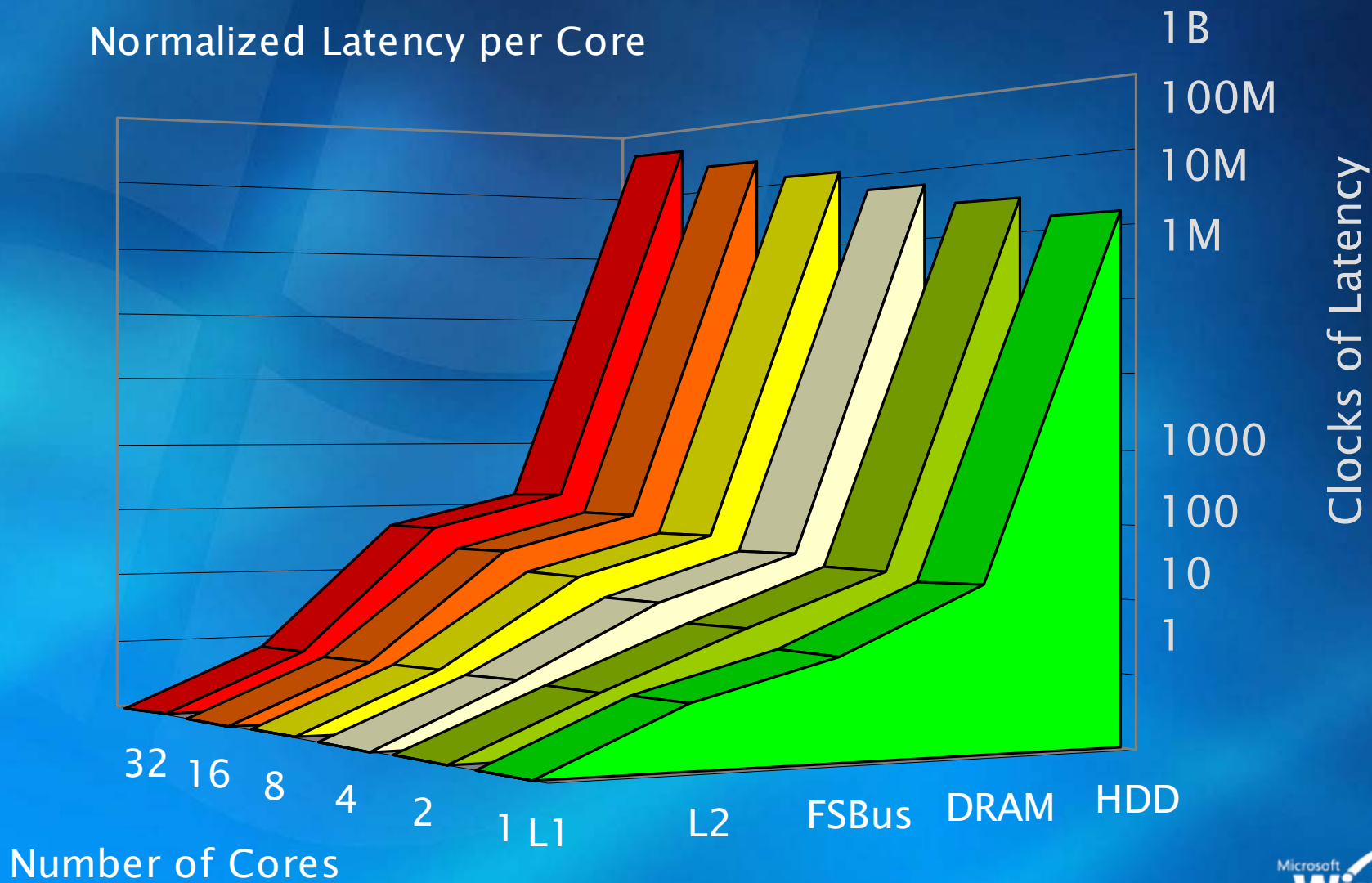
# Memory Trends

- Increasing density
- Faster interfaces
- Increasing latency
- NAND

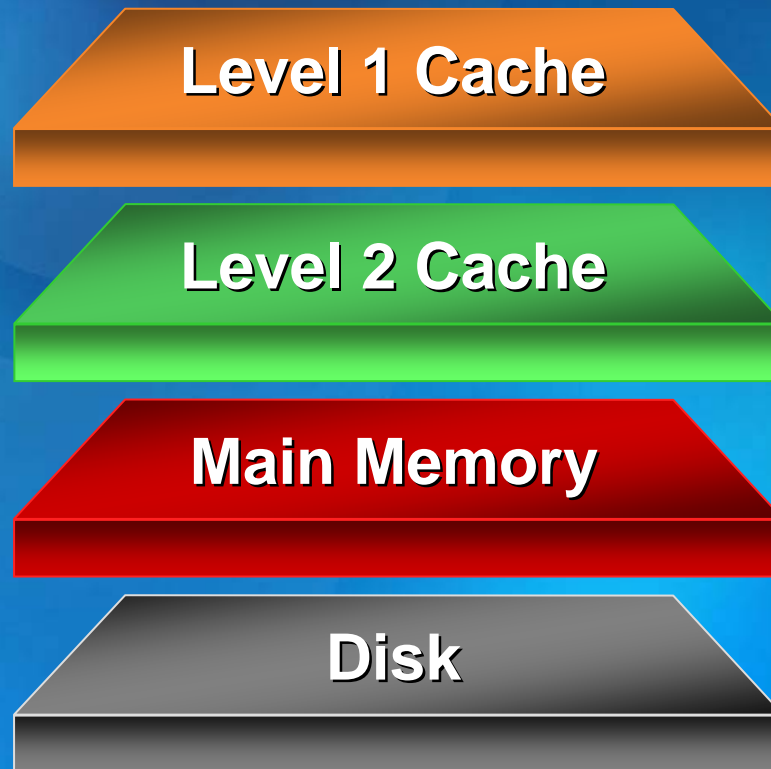


# Growing Gaps

Normalized Latency per Core

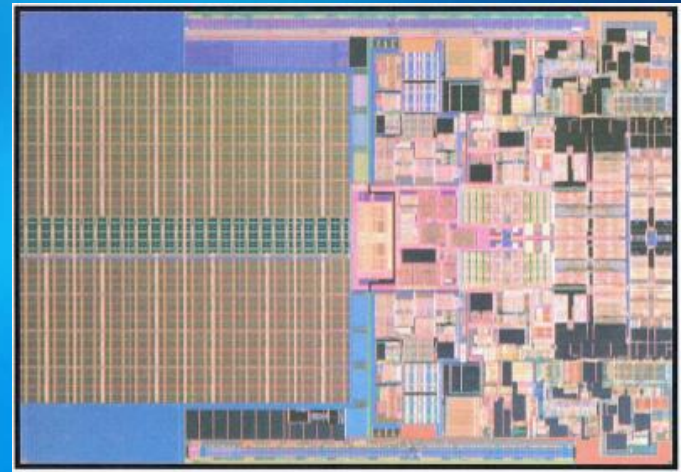
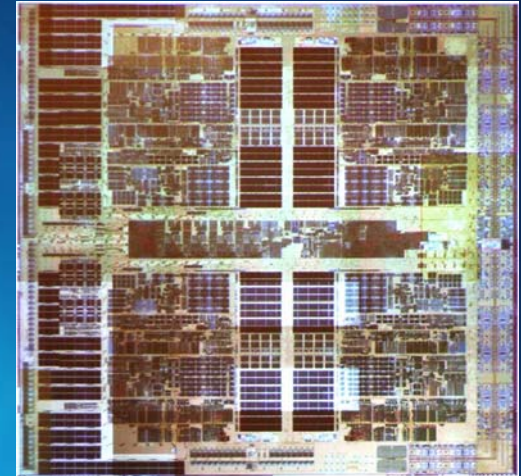


# Memory Hierarchy Expansion

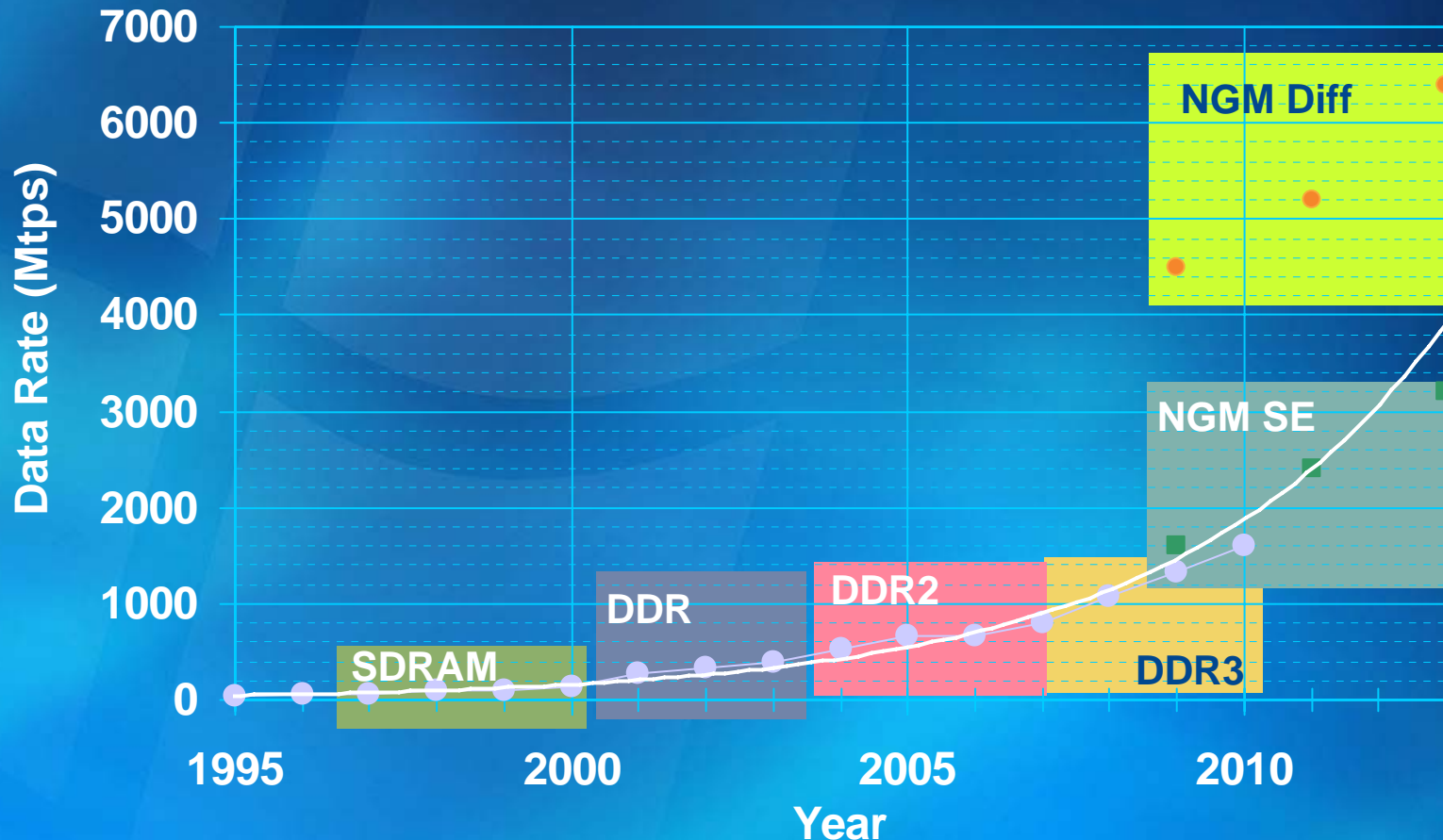


# Processor Trends

- AMD “Barcelona”
  - Quad core
  - 2M shared L3 cache
  - Dedicated L2 caches
- Intel “Penryn”
  - Dual/Quad core
  - 6MB/12MB L2 cache
- Intel “Nehalem”
  - Quad/8 core



# Main Memory Data Rate Trends



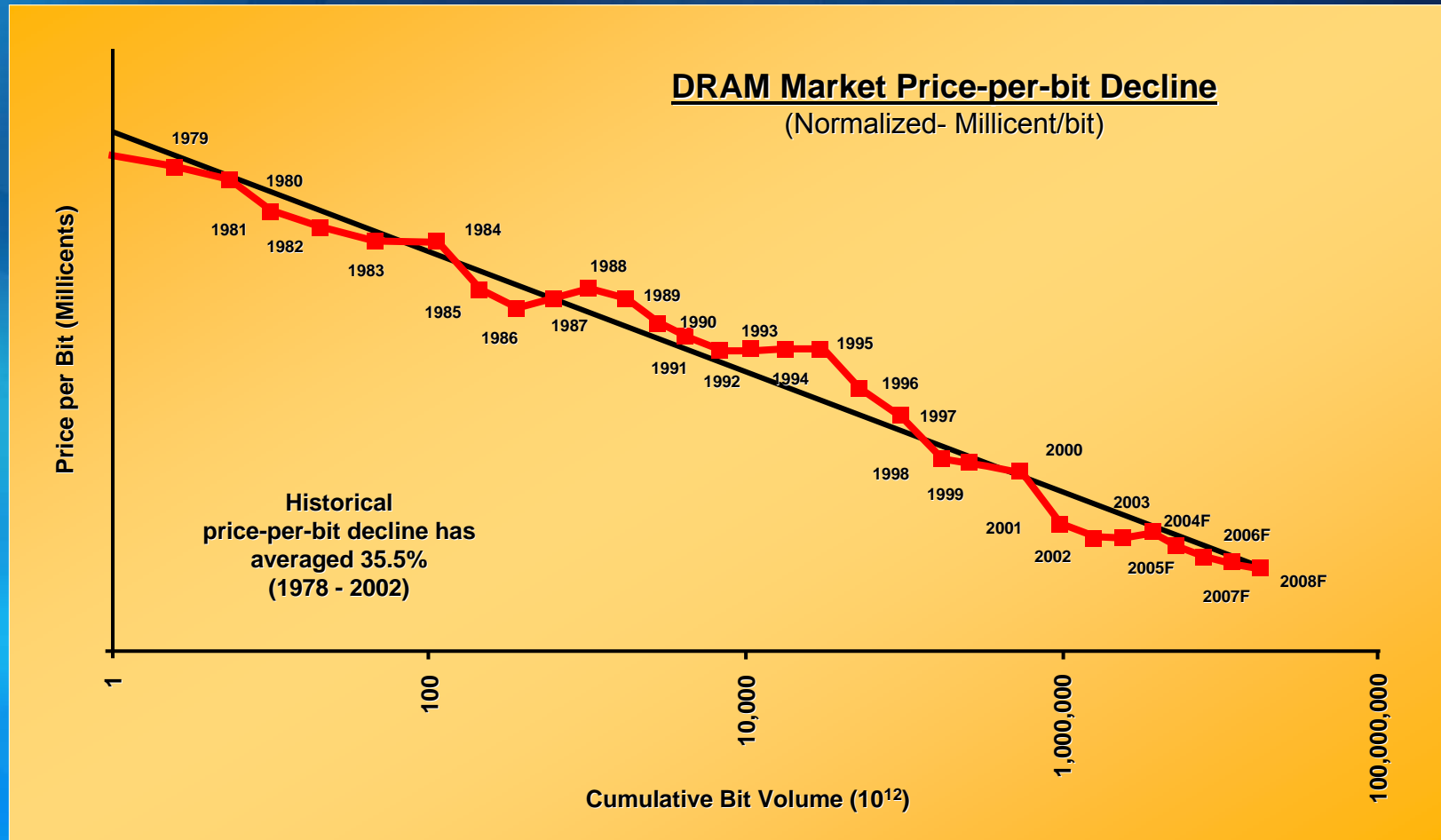
- DRAM bandwidth requirements typically double every 3 years

# Memory Trends

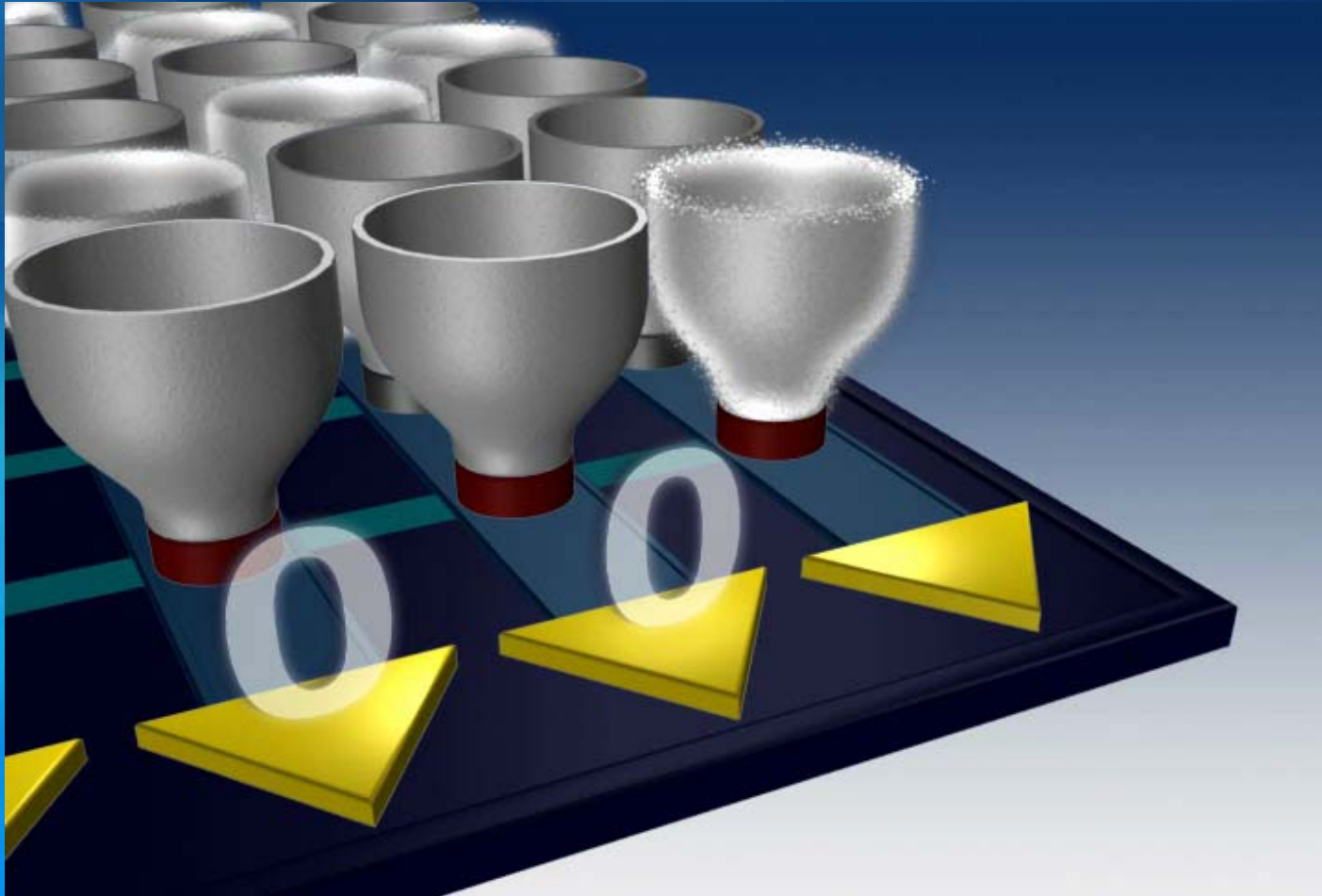
- But Latency is actually getting WORSE...
- And power is a problem...
- What drives memory evolution today?

## Economics and Physics

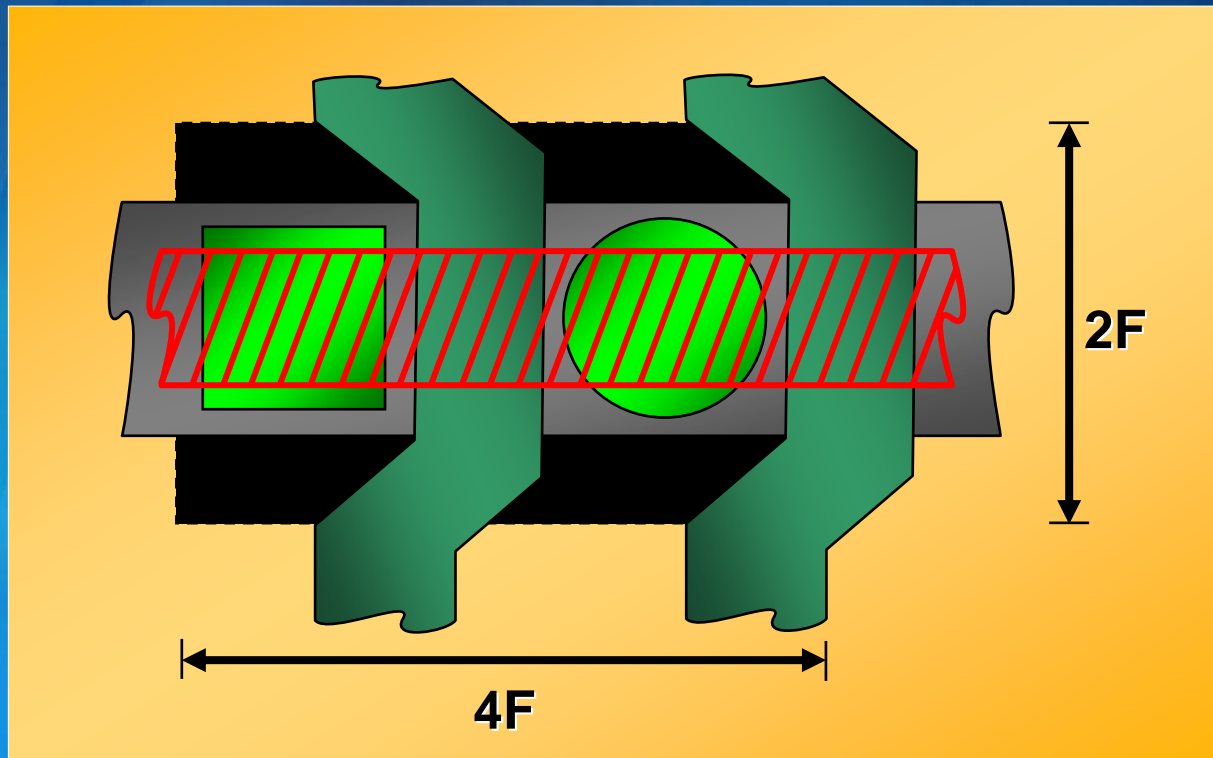
# Economics Drives Memory



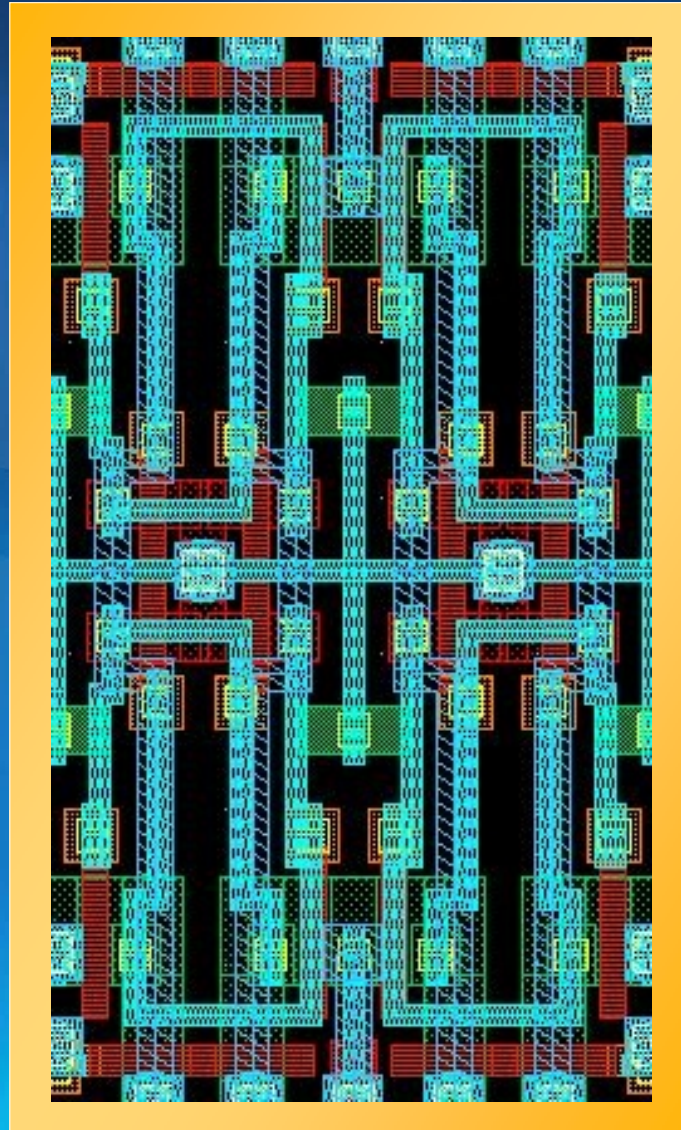
# Physics Drives Memory



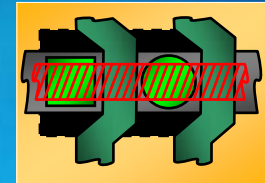
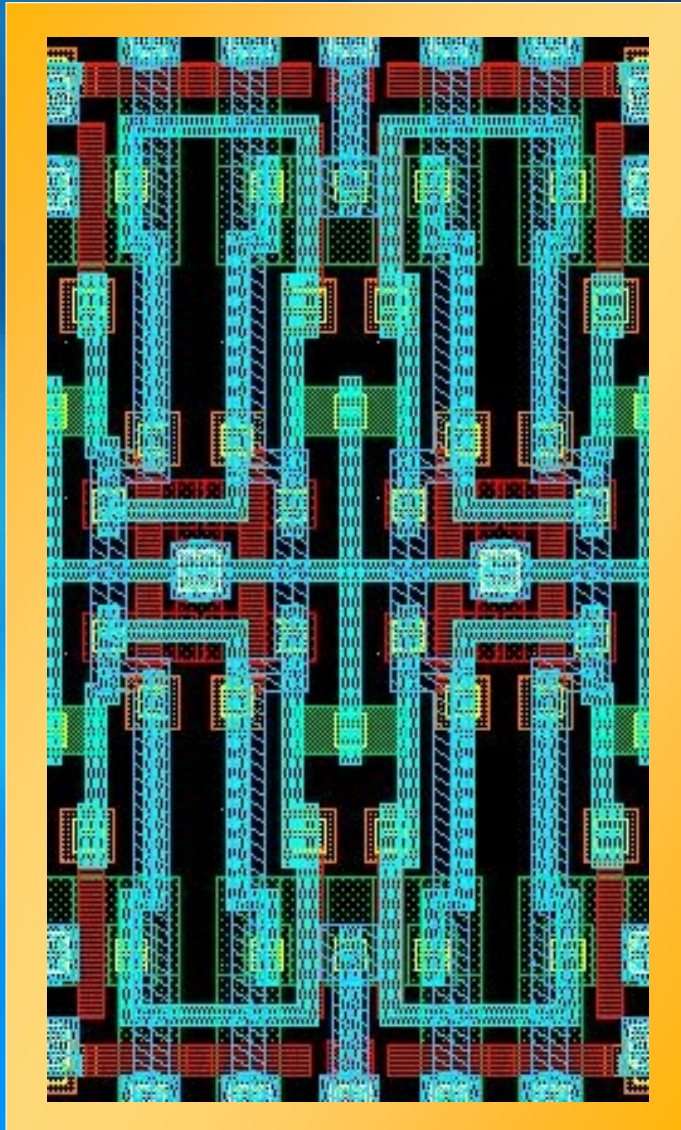
# DRAM Cell Layout: $8F^2$



# SRAM Cell Layout: 140F<sup>2</sup>



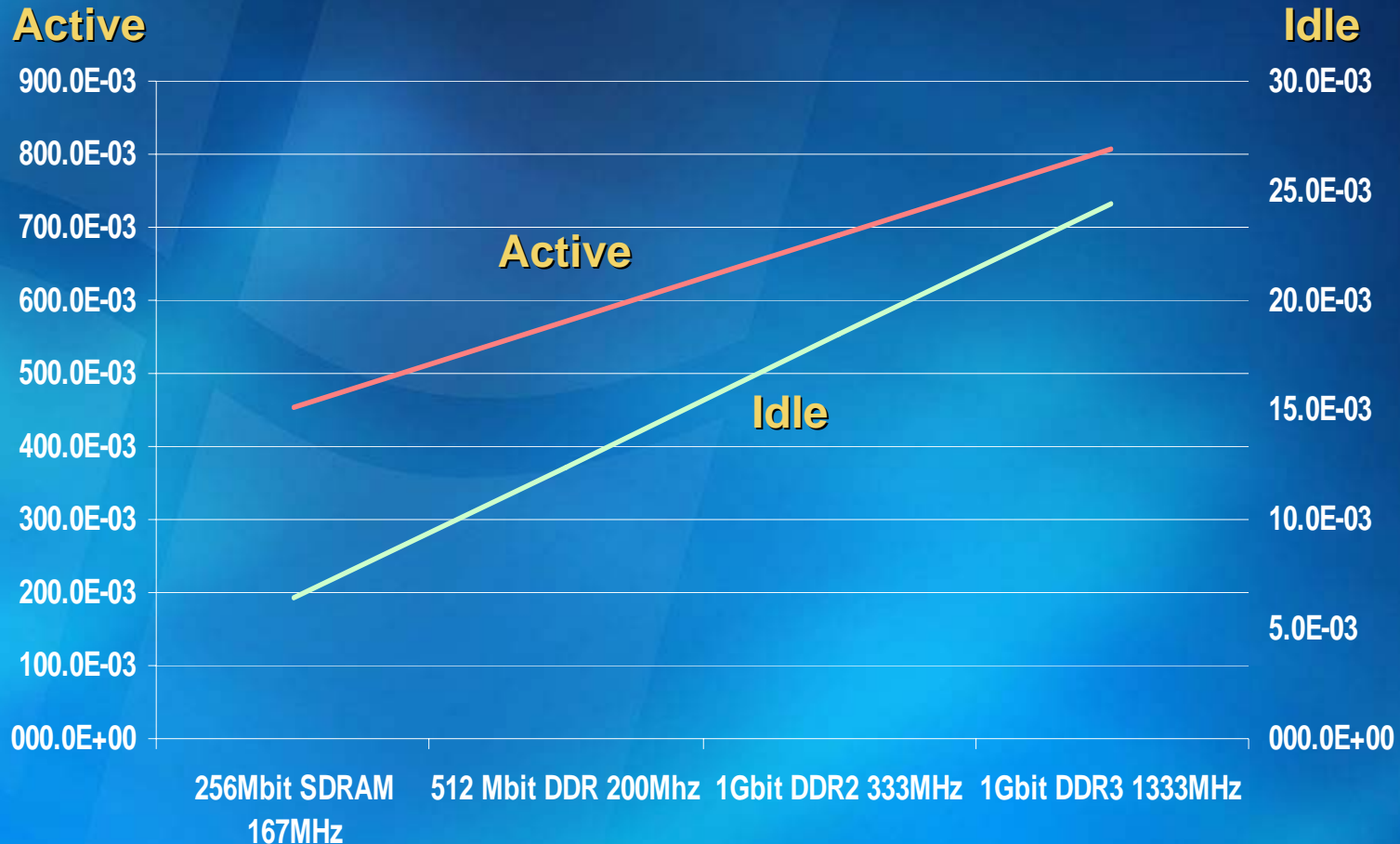
# Compared:



# Cell Sizes Compared

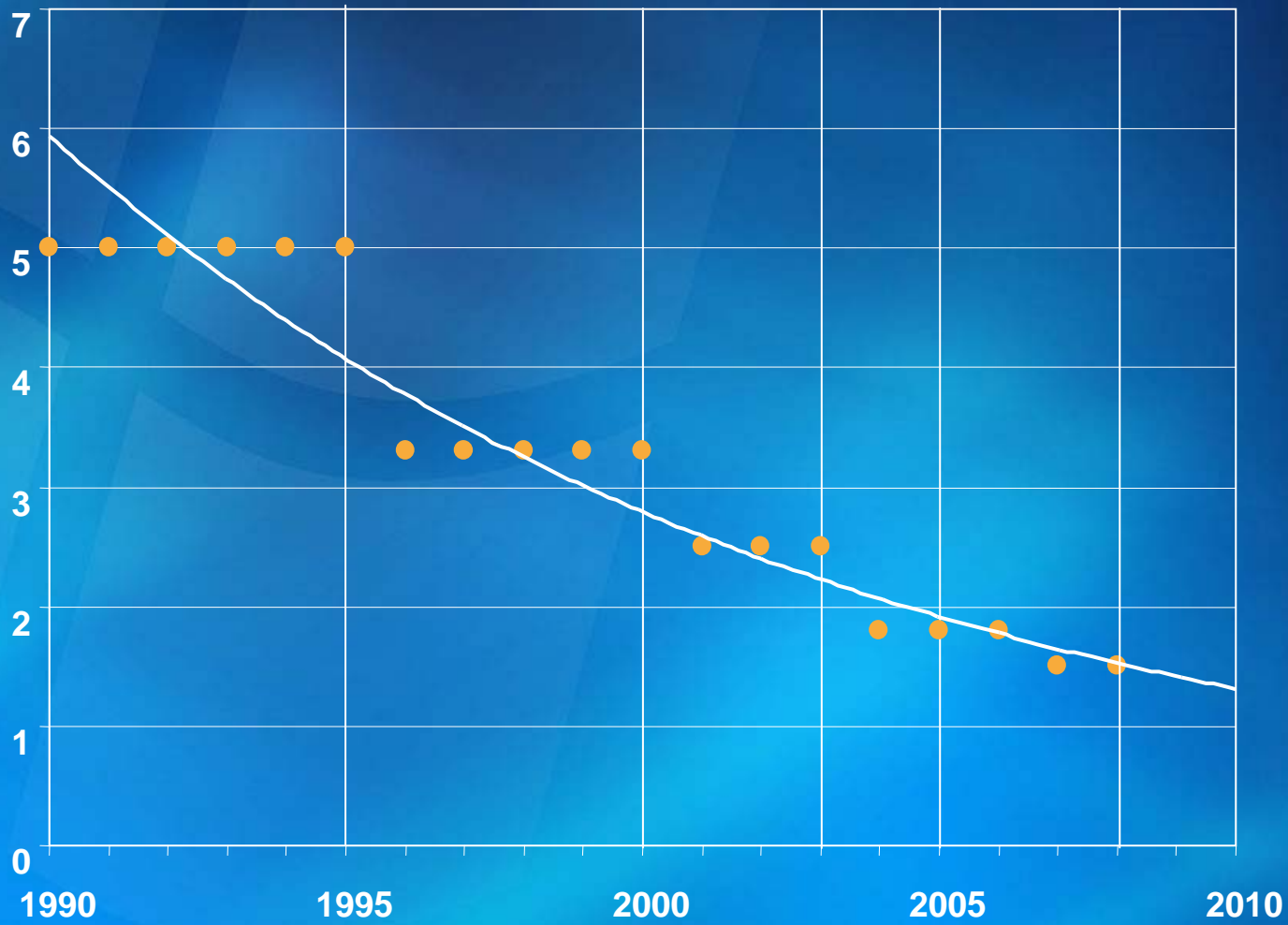
Cell Size ( $\mu^2$ )	Tech Node (nm)	Cell Size(F <sup>2</sup> )
IBM/Infineon MRAM		
1.42	180	44
Motorola 6T-SRAM		
1.15	90	142
0.69	65	163
Intel 65nm process 6T-SRAM		
0.57	65	135
Motorola eDRAM		
0.12	65	28
Motorola TFS: Nanocrystalline		
0.13	90	16
Micron 30-series DRAM		
0.054	95	6
Samsung 512Mbit PRAM Device		
0.050	95	5.5
Micron 50-series NAND		
0.013	53	4.5

# Power Trends

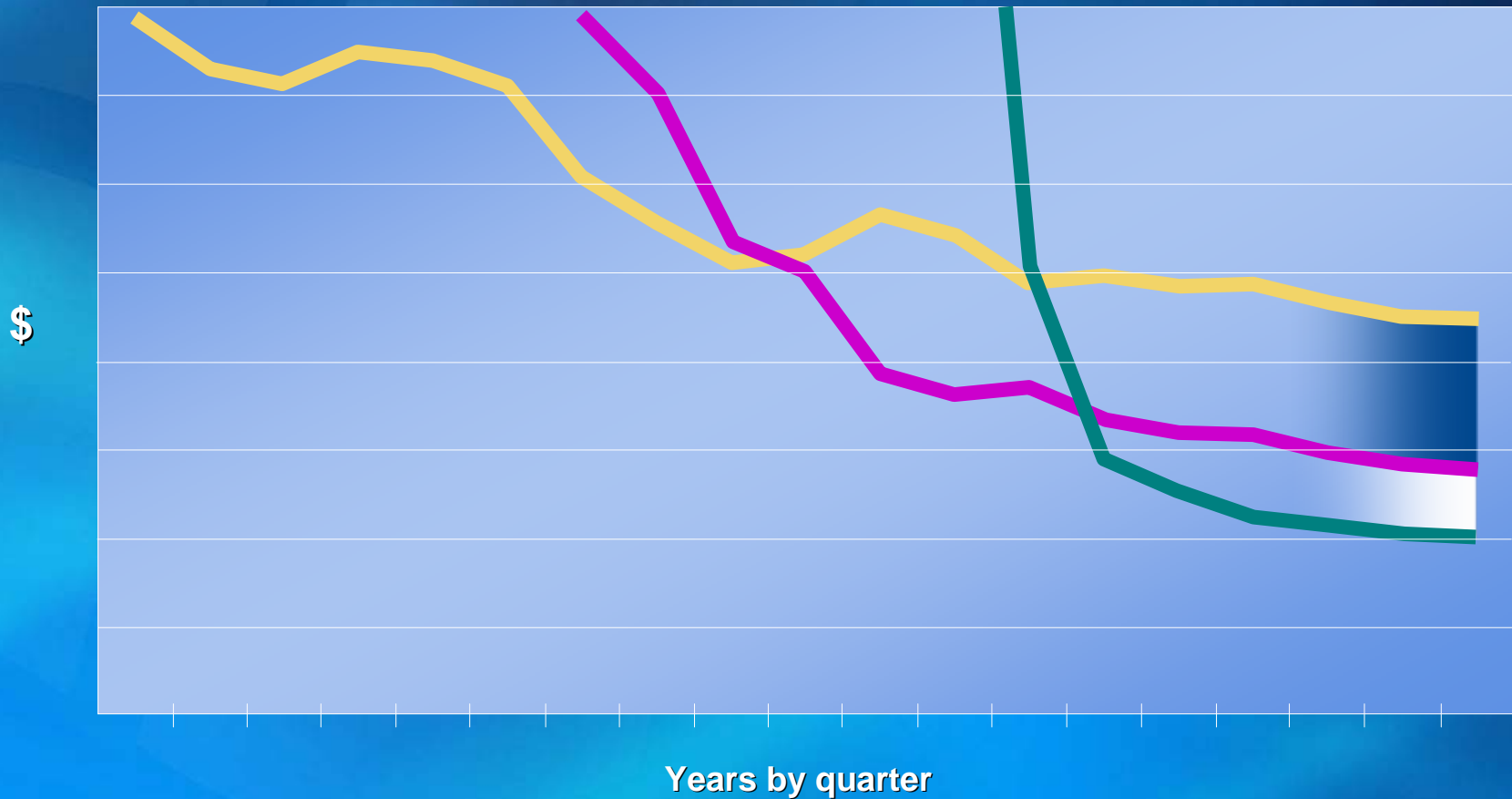


- X16 devices at nominal Vdd, linear trendlines

# Voltage Scaling Trends

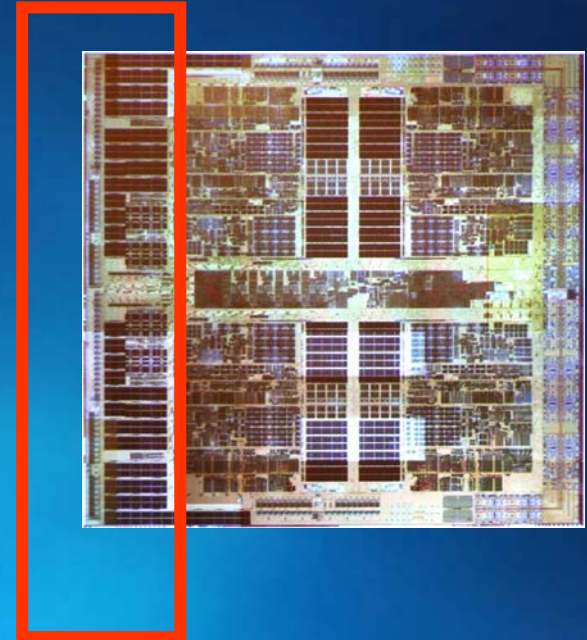
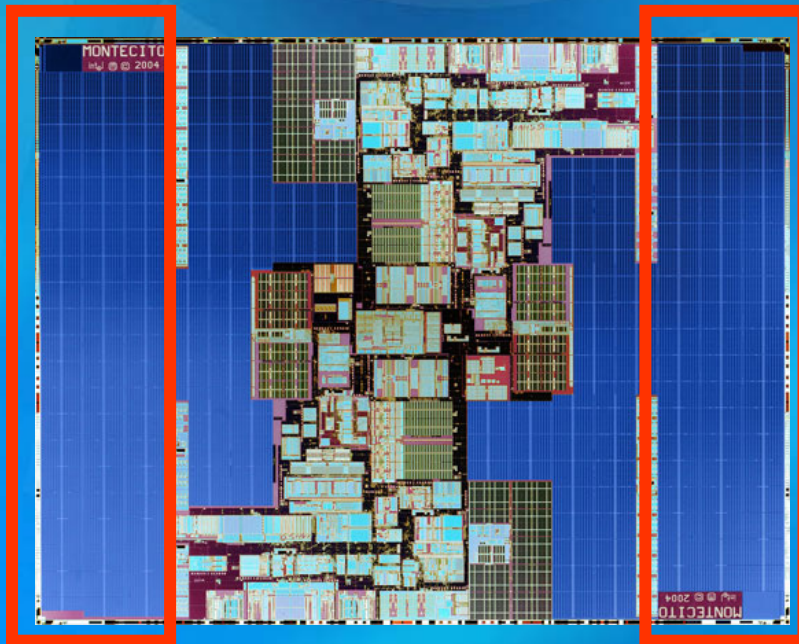


# Process Cost Increasing



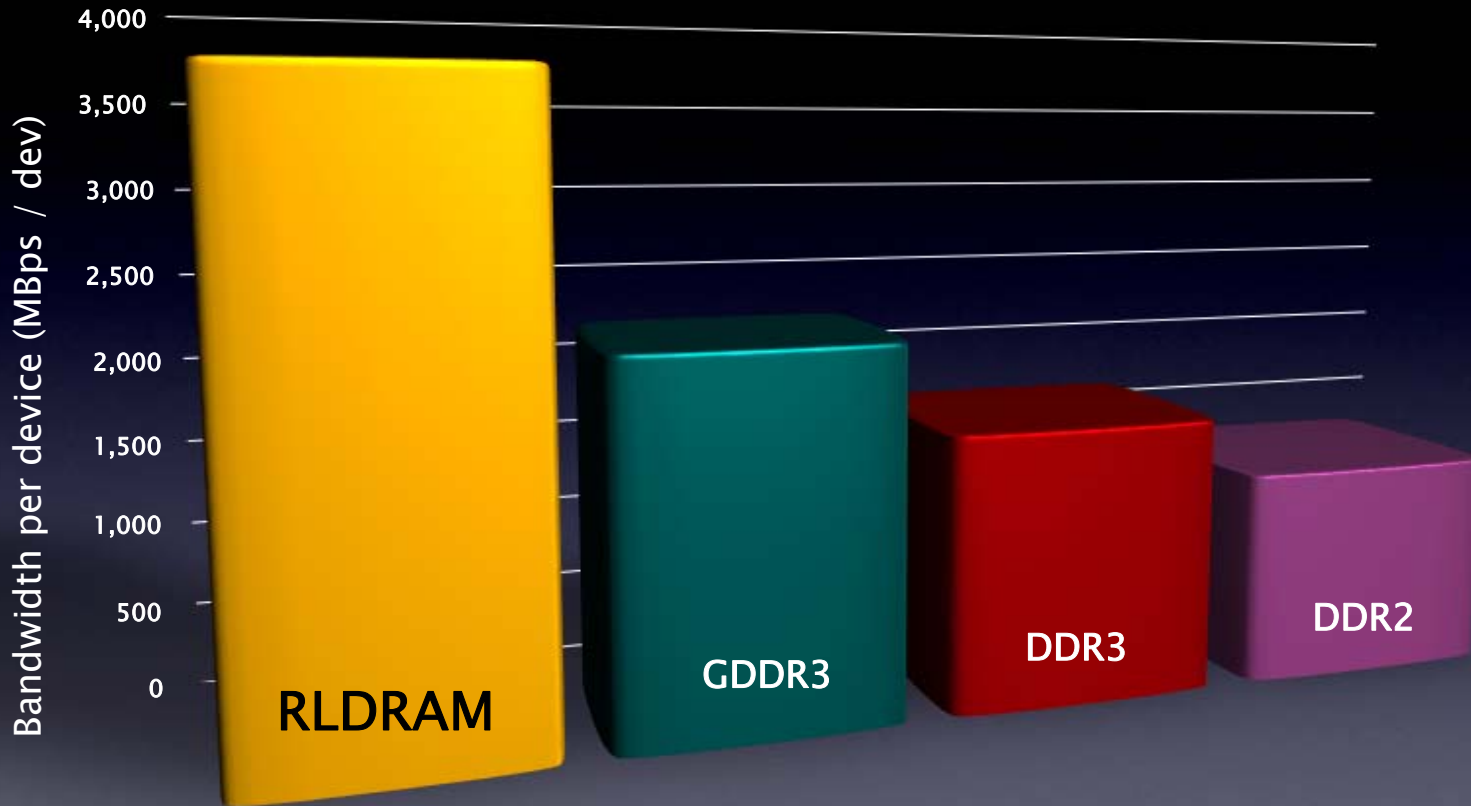
# Options for L3 Cache

- SRAM L3



# DRAM can be Fast

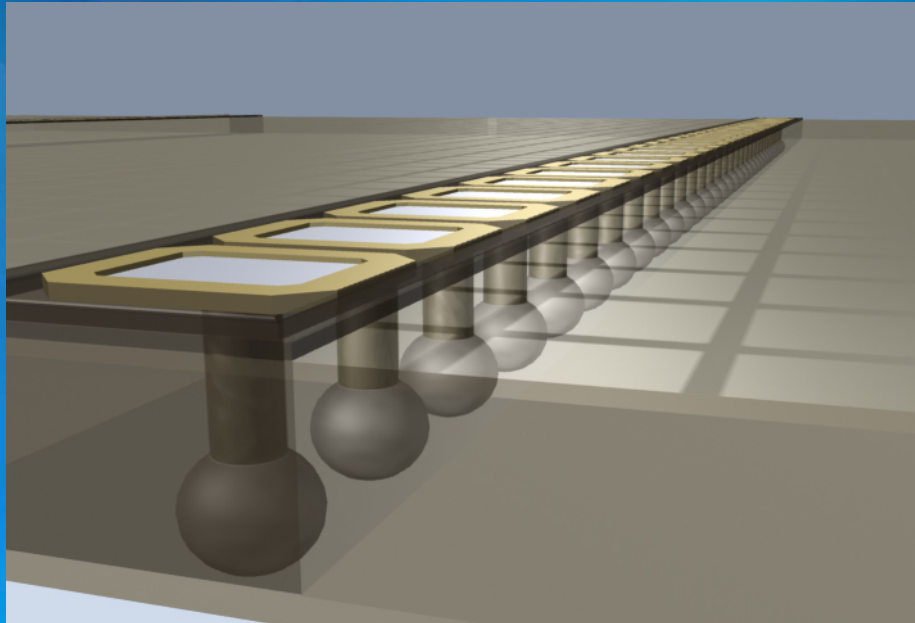
Random 16Byte Transfers Max Envelope



Access pattern: 8 READS followed by 8 WRITES

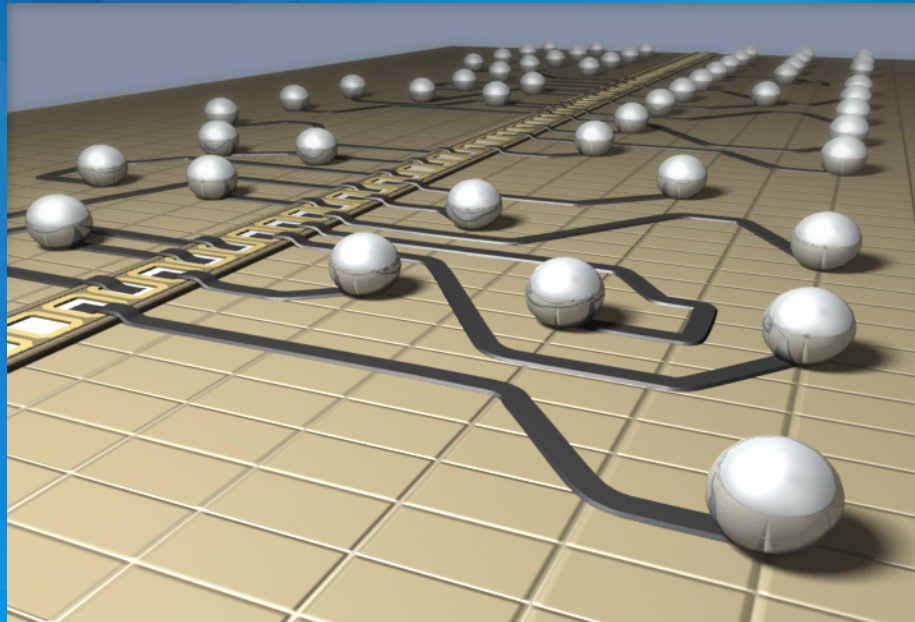
# Through-Wafer Interconnect

- Reduced parasitic loads
- Smaller ESD structures
- Greater numbers of interconnects
- TWI:



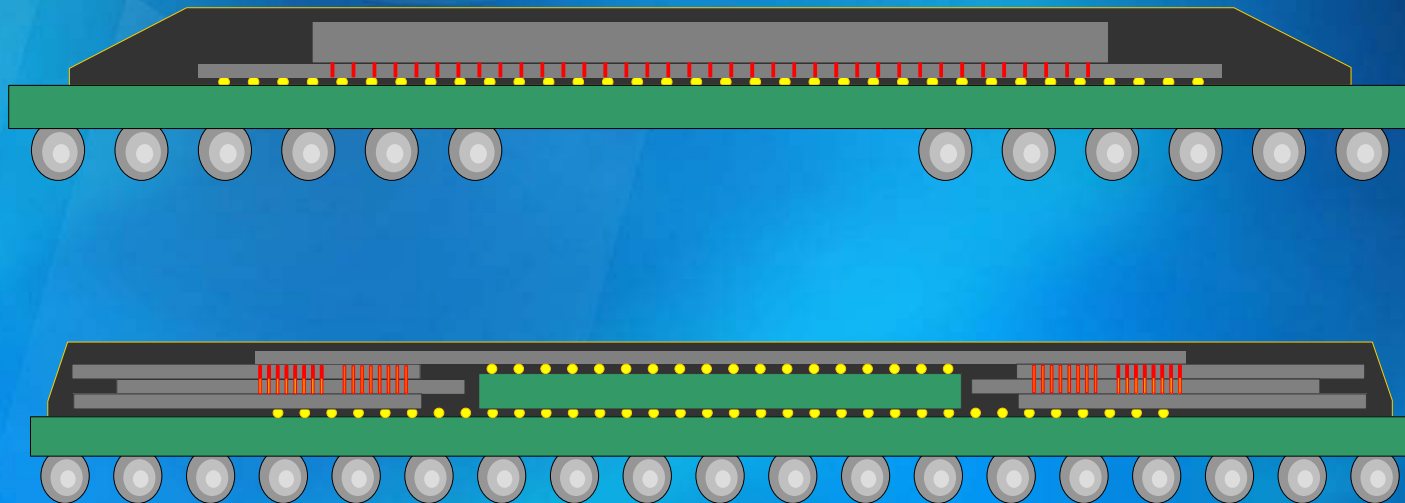
# Redistribution Layers

- Allow layout flexibility
- Reduced parasitic loads
- Supports great numbers of interconnects
- RDL:



# Stacked Silicon

- Goal of TWI and RDL
- Supports  $N \geq 2$  layers of silicon
- Supports processes optimized for device



# Storage Demand

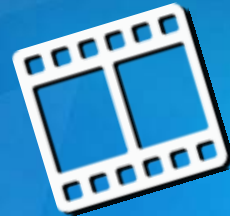
- 161 exabytes of digital data were generated in 2006

That's about 168 million terabytes, or roughly equivalent of:



1

million copies  
of every book  
in the Library  
of Congress



36

billion  
digital  
movies



43

trillion  
digital  
songs

# DRAM to Disk Evolution

- “Flash is Disk, Disk is Tape”?
- Performance, not capacity, is the issue.
- Disk will continue as the \$/bit leader
- NAND pricing is on a steep decline

# SSDs vs HDDs

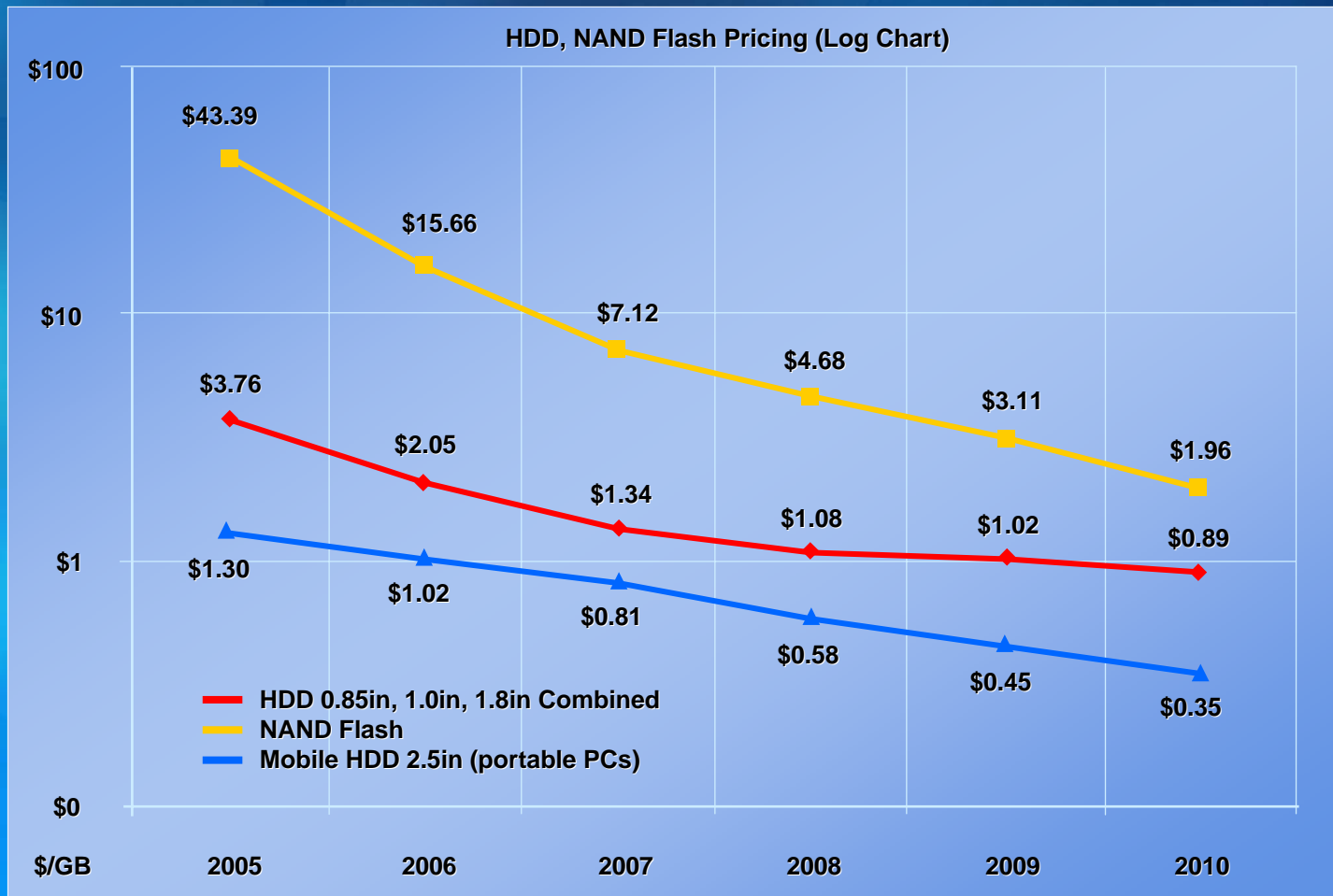
	SSD	HDD
Capacity		✓
Performance	✓	
Reliability	✓	
Endurance	✓	✓
Power	✓	
Size	✓	
Weight	✓	
Shock	✓	
Temperature	✓	
Cost per bit		✓

- Based on recent advances in NAND lithography, SSD densities reach capacities for mass market appeal
- SSD offers many features that lead to improved user experiences
- Early shortcomings for reliability and endurance have been overcome

NAND Solid State Storage Devices are ready for deployment in many applications

# NAND Density Trends

## ● Beating Moore's Law



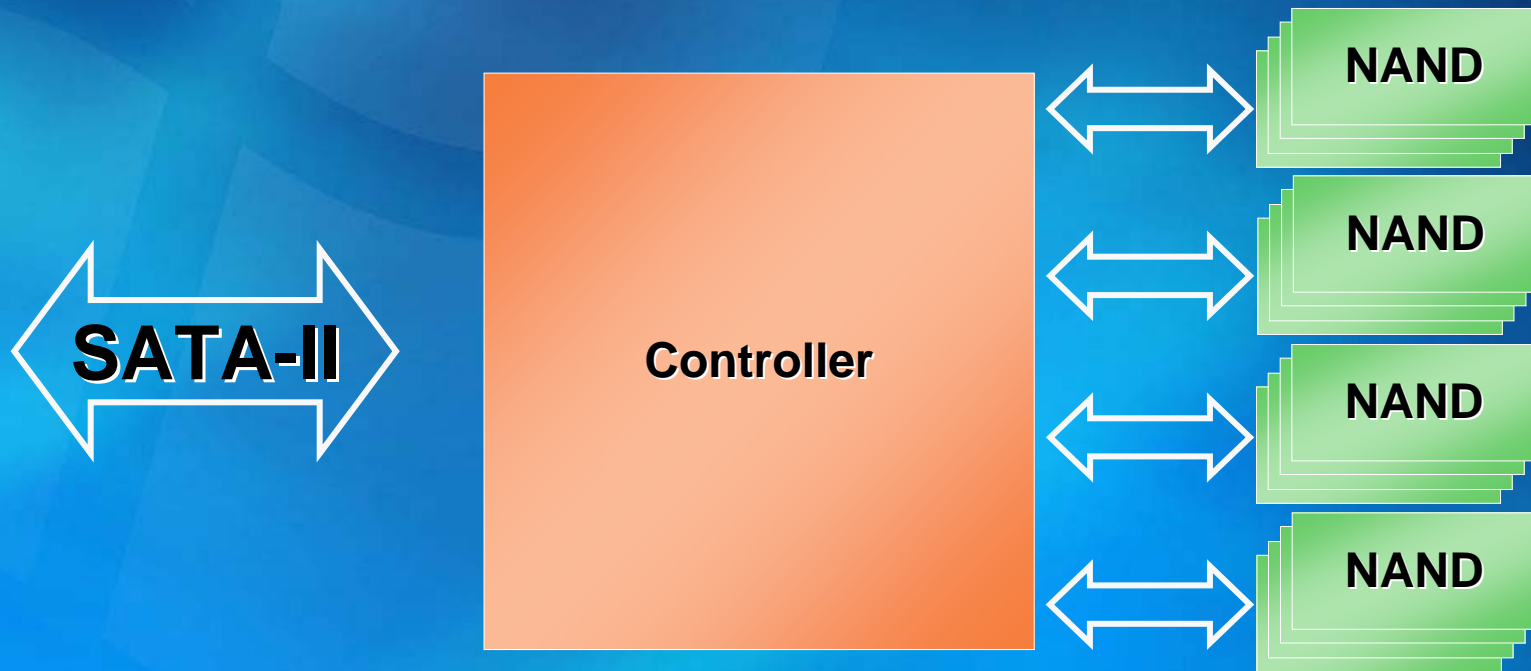
# NAND in Notebooks and Consumer

Average Specifications	Hard Disk Drive	Solid State Drive	Hard Disk Drive	Hybrid Hard Drive
	1.8" HDD	SSD(1.8"/2.5")	2.5" HDD	2.5" HHD
Capacity	30-80 GB	4-32GB	40-160GB	Up to 160GB
Data Rate (max sustain)				
Read	25MB/s	57MB/s	44MB/s	
Write	25MB/s	32MB/s	44MB/s	
Spindle Speed	4200 RPM	None	5400 RPM	5400 RPM
Seek	15 ms	None	12 ms	12.5 ms
Non op shock	1500 G	2000 G	900 G	900 G

- **SSD and HHD both provide power savings in various applications, but the exact power savings fluctuate from application to application**
- **In a test of a 32GB SSD drive, the power savings of the SSD was 1 watt better than the closest tested HDD**

# What to Look for in an SSD

- SSD-optimized controller
- Parallel NAND channels



# SSDs in the Enterprise



NAND

CPU		
Relative Latency		Relative Cost/bit
1	L1 Cache	200
2.5	L2 Cache	140
35	L3 Cache	120
300	DRAM	8
250,000	SSD	3
25,000,000	HDD	0.7

**NAND Flash Closes the Latency Gap**

Cost/bit Data as of Aug '06

# Datacenter Issues

- Power
- Reliability
- Space
- Performance

Table 1: Observed failure rates

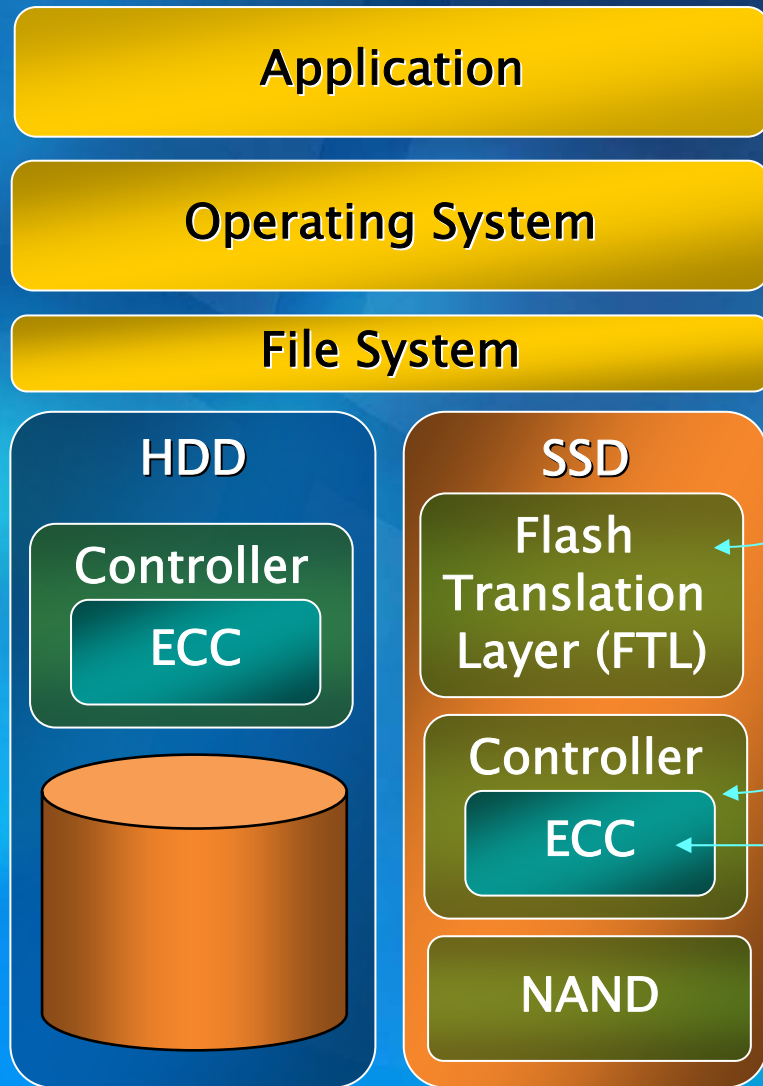
System	Source	Type	Part Years	Fails	Fails /Year
TerraServer SAN	Barclay	SCSI 10krpm	858	24	2.8%
		controllers	72	2	2.8%
		san switch	9	1	11.1%
TerraServer Brick	Barclay	SATA 7krpm	138	10	7.2%
Web Property 1	anon	SCSI 10krpm	15,805	972	6.0%
		Controllers	900	139	15.4%
Web Property 2	anon	PATA 7krpm	22,400	740	3.3%
		motherboard	3,769	66	1.7%

# Reliability and Endurance

	Effect	Description	Observed as...	Management
Reliability	Program Disturb	Cells not being programmed receive charge via elevated voltage stress	Increased read errors immediately after programming	ECC and Block Management
	Read Disturb	Cells not being read receive charge via elevated voltage stress	Increased read error at high number of reads	ECC and Block Management
	Data Retention	Charge loss over time	Increased read errors with time	ECC and Block Management
Endurance	Endurance/ Cycling	Cycles cause charge trapped in dielectric	Failed Program/Erase Status	Retire Block

NAND failure mechanisms are well understood and managed

# Management Stack



## Flash Translation Layer

- Interfaces to traditional HDD File System
- Enables sector I/O to Flash
- Wear leveling
- Bad block management
- Automatic reclamation of erased blocks
- Power loss protection
- Manages multiple NAND devices

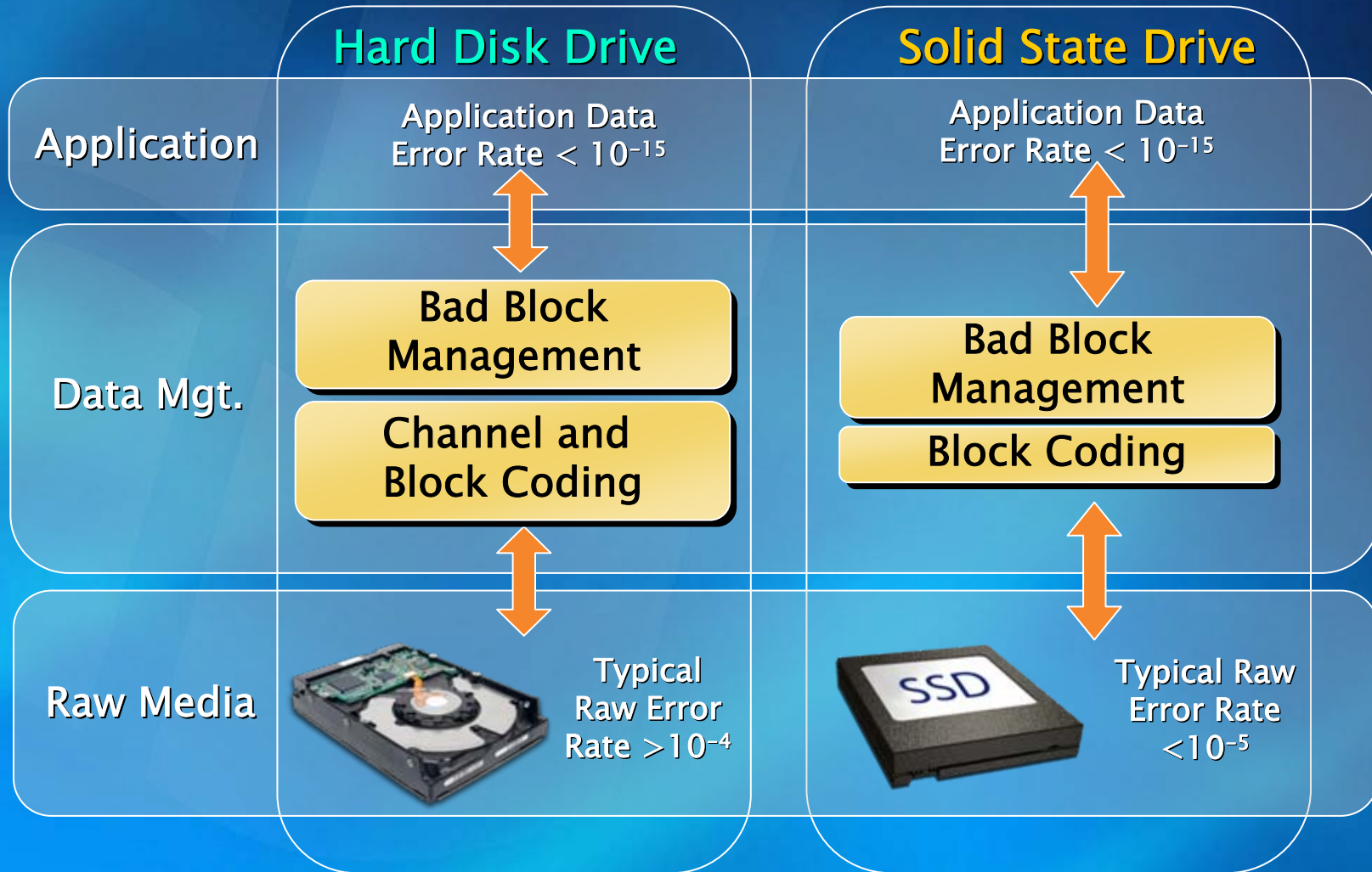
## Controller

- Manages Physical Protocol
- NAND Command encoding
- High Speed data transport (DMA/FIFO)

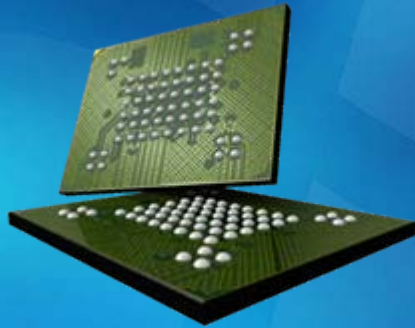
## Error Control

- Algorithm to control sector-level bit reliability
- Implemented in hardware with software control
- Algorithm depends upon Flash technology

# SSD & HDD Reliability



# SSD Quality and Reliability



## NAND Specs

- 100K P/E Cycles
- 1-Bit ECC
- Limited Read Cycles
- 10 year data retention



## NAND

- Extended operation of NAND
- On-going production management to assure reliability

## Management

- NAND-Validated Error Correction
- Static and dynamic Wear Leveling
- Garbage collection
- Bad block remapping
- Other proprietary schemes

## Applications

- Optimizations based upon Management and NAND



## SSD Specs

- 10 Year operating life
- $10^{-15}$  Bit Error Rate
- 1E+6 hours MTBF

# Endurance: Usage Example



Capacity: 32 GB

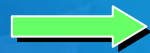
Continuous write at max bus speed of 100MB/s with a 5:1 r/w ratio



30 minutes  
to fill up  
disk



50 years before 1M  
I/O cycle limit is  
exceeded



At 20%  
utilization



250 years before 1M I/O  
cycle limit is exceeded

Opportunities for  
improvement, i.e. new  
coding, will further extend  
time to cycle limit

Micron Flash Drives are ready for deployment for various applications

# Wear Leveling

- Wear leveling is a plus on SLC devices where blocks can support up to 100,000 PROGRAM/ERASE cycles
- Wear leveling is imperative on MLC devices where blocks can typically support less than 10,000 cycles
- If you erased and reprogrammed a block every minute, you would exceed the 10,000 cycling limit in just 7 days!

$$60 \times 24 \times 7 = 10,080$$

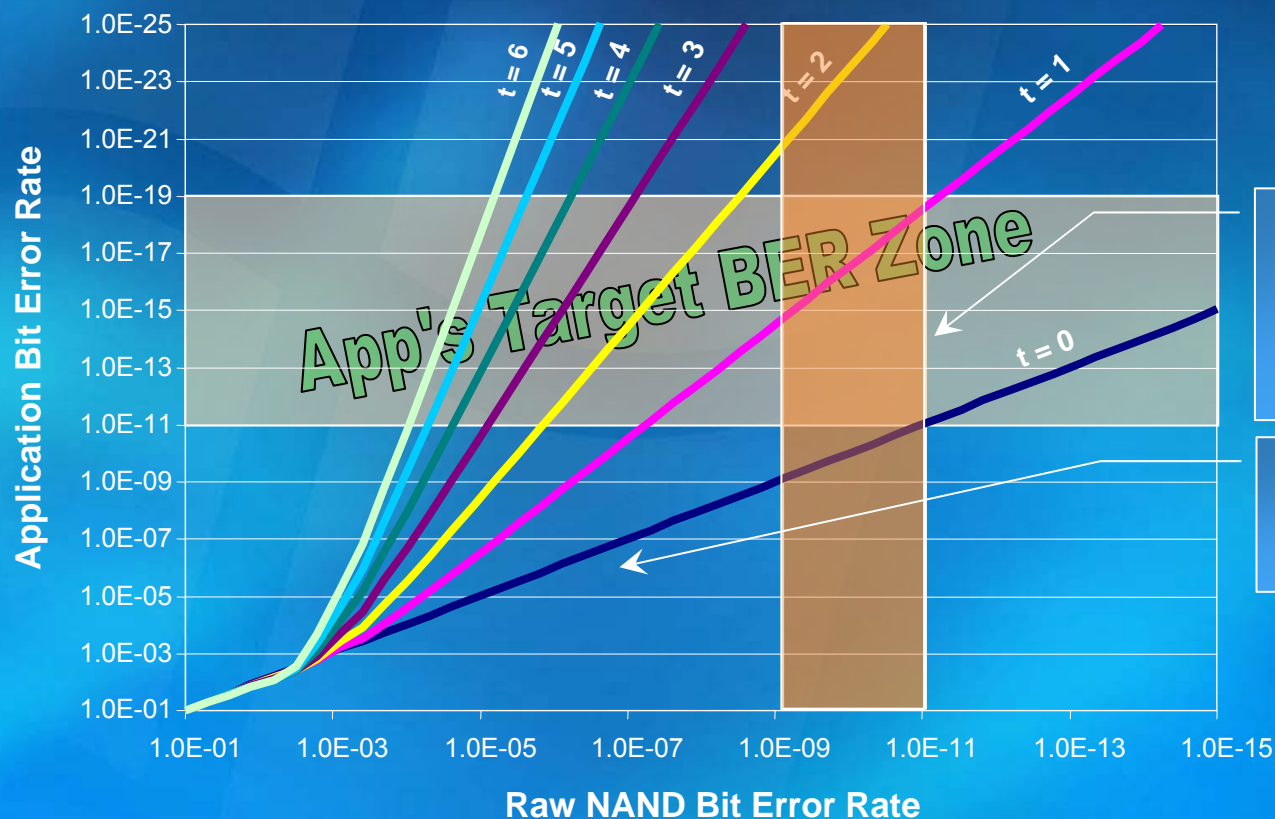
- Rather than cycling the same block, wear leveling involves distributing the number of blocks that are cycled

# Wear Leveling (continued)

- An 8Gb MLC device contains 4,096 independent blocks
- If we took the previous example and distributed the cycles over all 4,096 blocks, each block would have been programmed less than 3 times (vs. the 10,800 cycles when you cycle the same block)
- If you provided perfect wear leveling on a 4,096 block device, you could erase and program a block every minute, every day for 77 years!

$$\frac{10,000 \times 4,096}{60 \times 24} = \frac{40,960,000}{1,440} = 28,444 \text{ days} = 77.9 \text{ Years}$$

# ECC Code Selection Becoming More Important



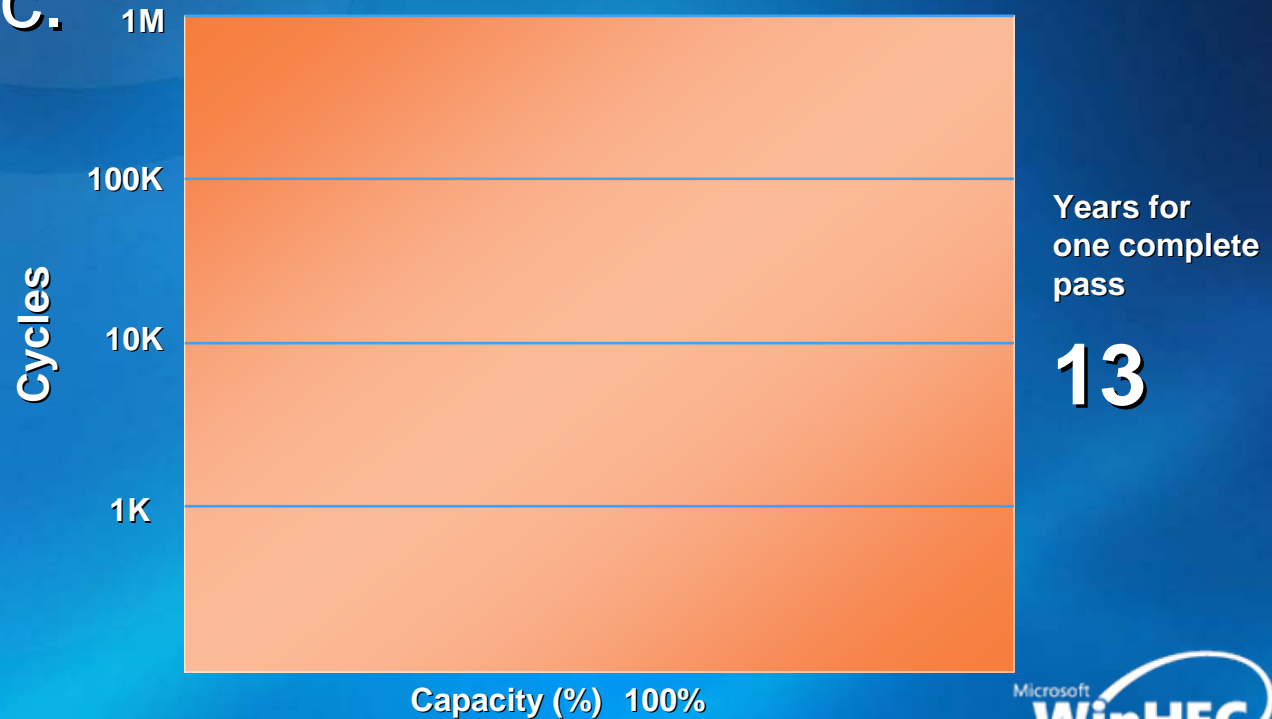
For SLC  
A code with correction  
threshold of 1 is sufficient

t = 4 required (as a  
minimum) for MLC

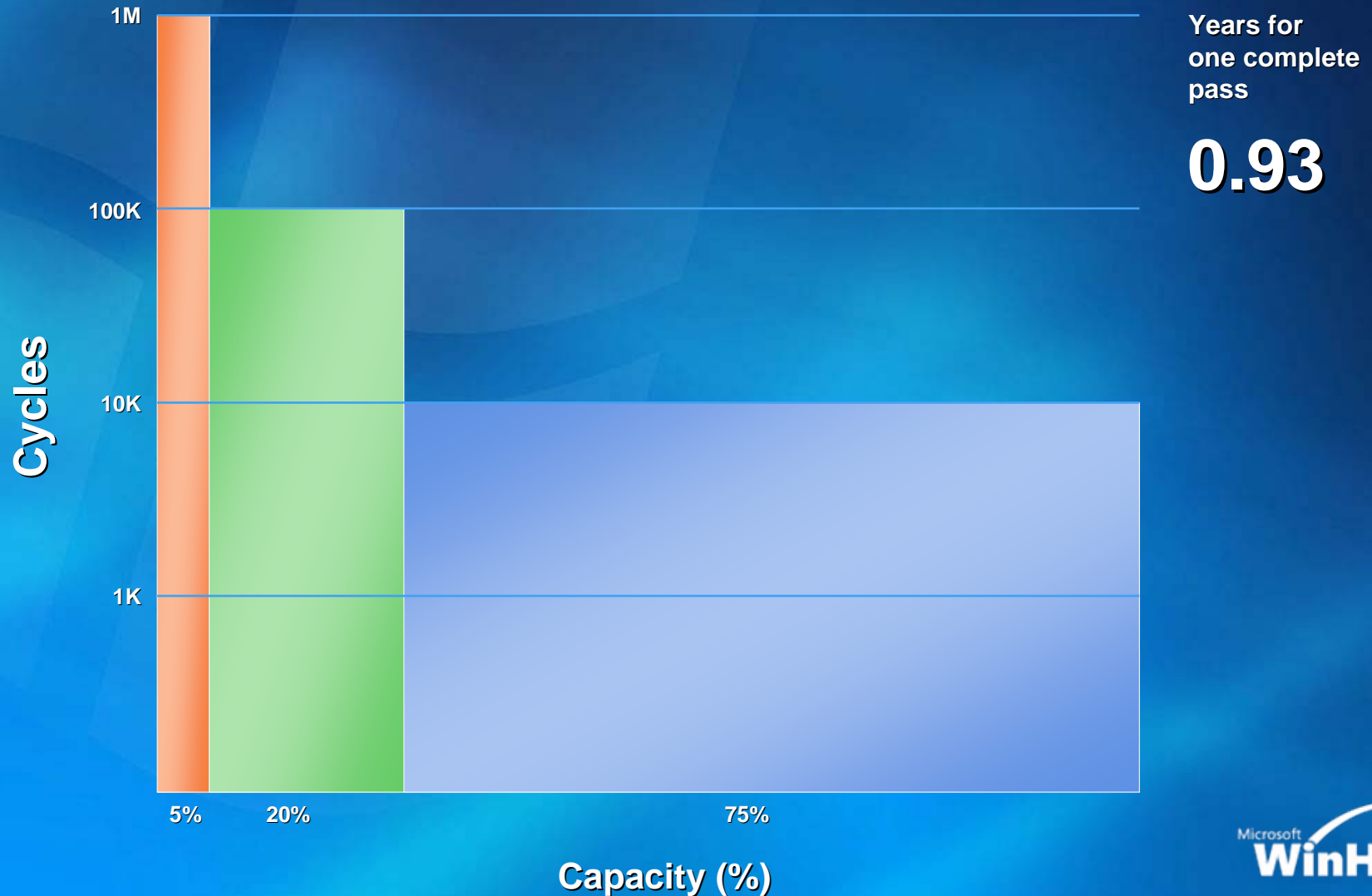
As the raw NAND Flash BER increases, matching the ECC to the application's target BER becomes more important

# Meaningful Cycling Metrics

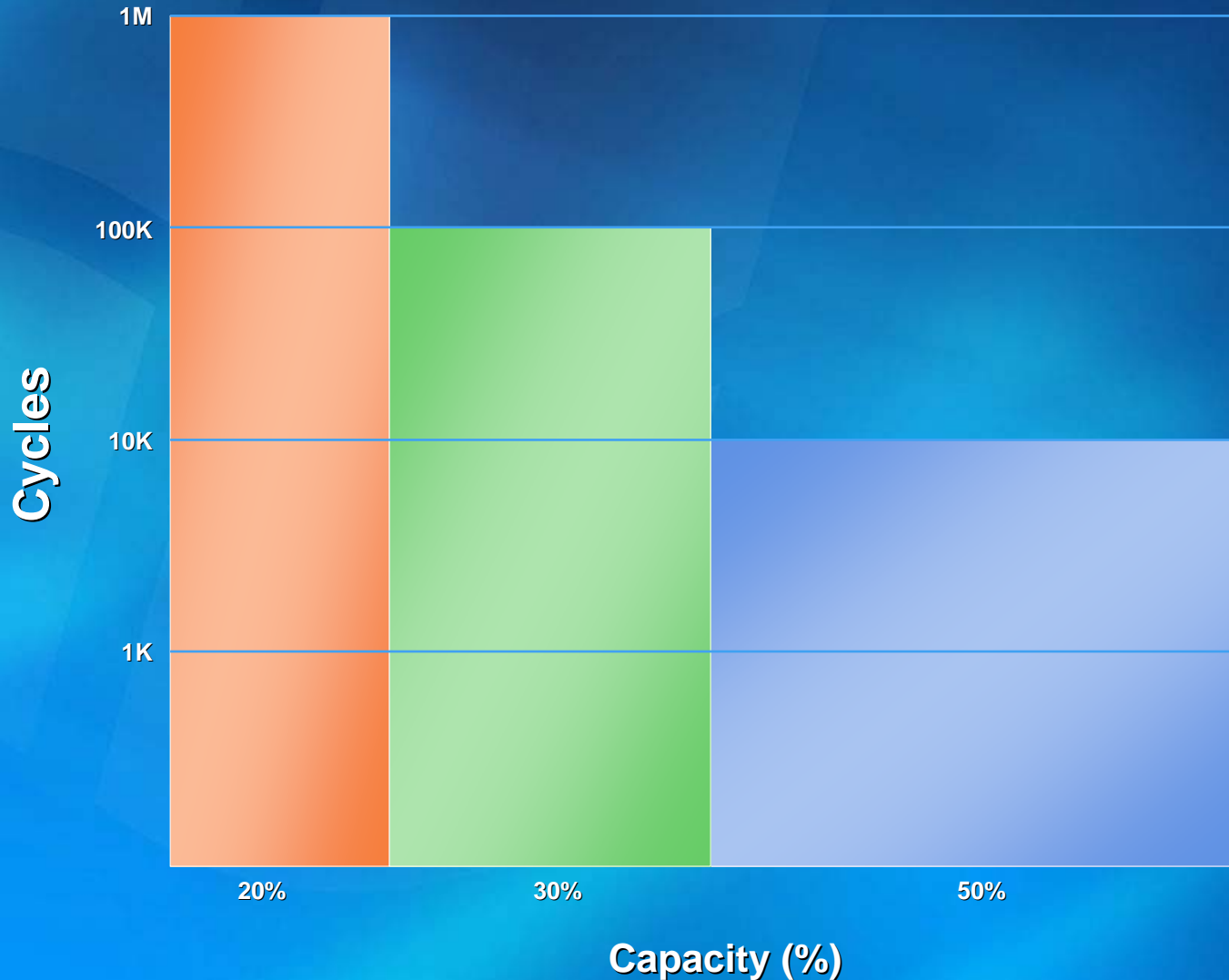
- Practical, testable solutions are needed.
- Simply stating that “the drive must meet 1 million complete read and write cycles” is not realistic.



# Cycling for CE Applications



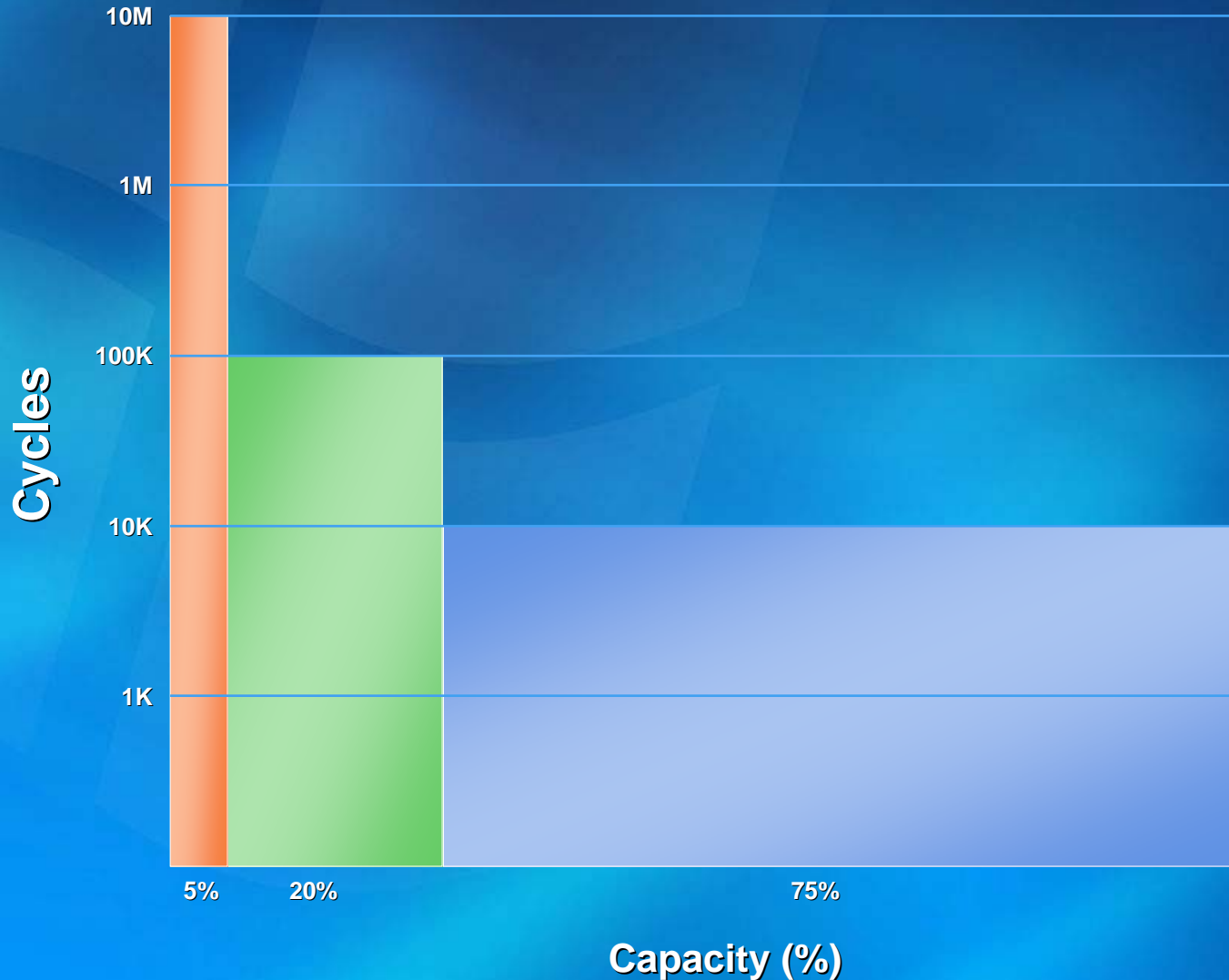
# Cycling for Servers



Years for  
one complete  
pass

**3.02**

# Cycling for Enterprise



Years for  
one complete  
pass

**6.78**

# Call to Action

- Close the Gaps!
- Innovation opportunities exist close to the CPU with DRAM-based caches.
- Innovation opportunities enabled by rapid NAND scaling for NAND-based storage.

# Additional Resources

- Web Resources:
  - Specs: <http://www.micron.com/winhec07>
  - Whitepapers: <http://www.micron.com/winhec07>
  - Related Sessions
    - Main Memory Technology Direction
    - Flash Memory Technology Direction
- Email address:
  - [daklein@micron.com](mailto:daklein@micron.com)

# **Microsoft<sup>®</sup>**

*Your potential. Our passion.<sup>™</sup>*