

"What will have the greatest impact in 2010:
The processor, the memory, or the interconnect?"

T. Basil Smith
IBM Watson Research Center
June 30, 2001

not to bias the discussion, but:

Its the memory, stupid!

Message

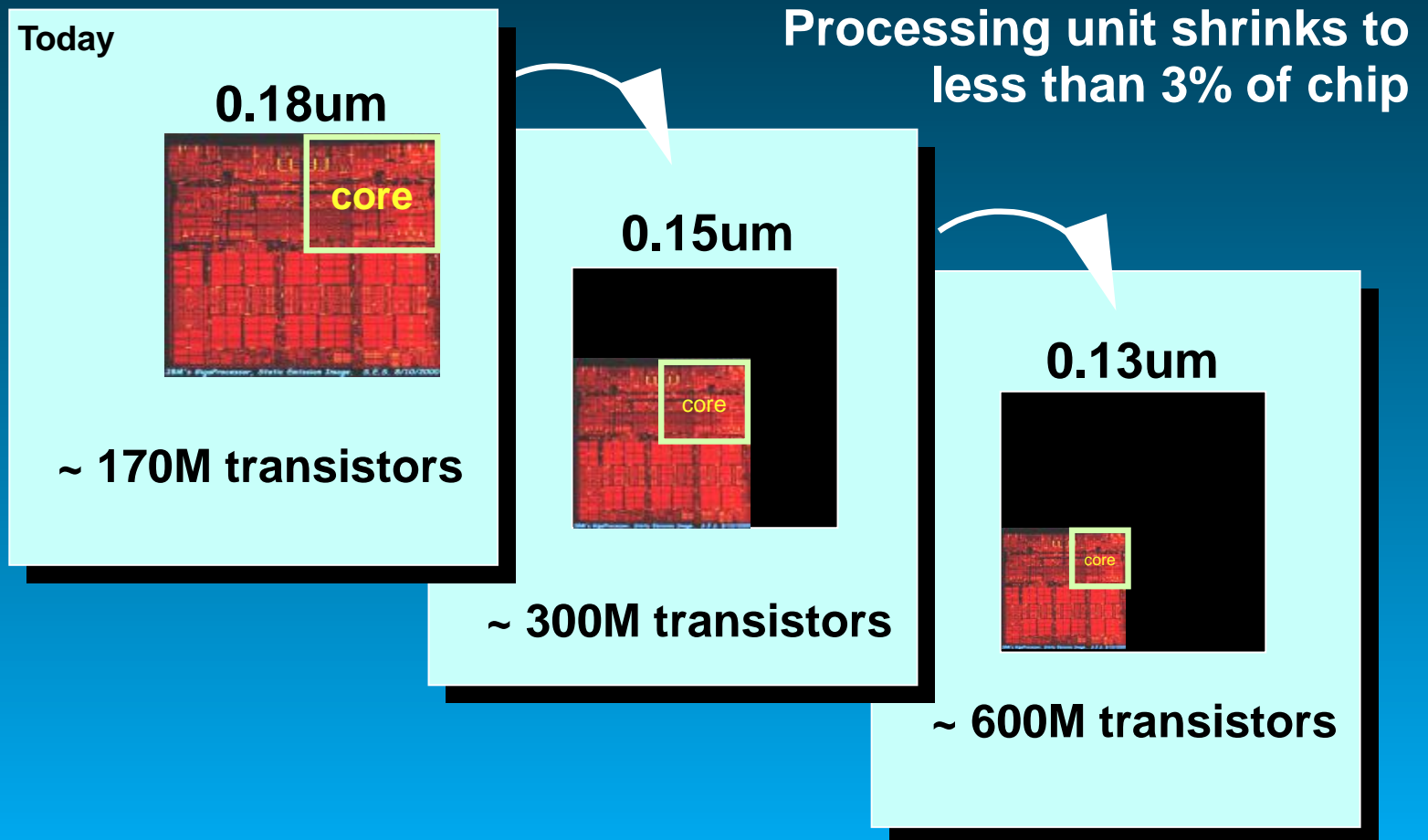
- **Memory technology is the most slowing improving of the critical systems technologies**
 - Processor technology (cycle time, CPI, SMP size, \$/MIPS)
 - Storage technology (\$/MB, density, bandwidth)
 - Interconnect technology (bandwidth)
 - Memory technology (latency, bandwidth, \$/MB)
- **Memory becoming the performance and cost bottleneck**
 - Improvements in system performance will be limited by memory hierarchy and memory characteristics.
 - Will overtake storage as the dominant cost component in systems
 - Should also displace storage as the first tier of stable storage

Summary

Years	2001	2004	2010
Attributes			
Processor Freq.	1-1.5GHz	3-5GHz	10-15GHz
On Chip Caches	256KB-2MB	1MB-8MB	16MB-128MB
Off Chip Caches	16MB-128MB	64MB-512MB	1GB-8GB
DRAM Latency (Mem Sys Lat.)	40-60ns (200-300ns)	40-60ns (200-300ns)	30-50ns (300-500ns)
Memory BW	3.2 GB/s	8.5 GB/s	40 GB/s
IO Latency	~2-10 ms	~2-10 ms	~2-10 ms
IO Bandwidth	1 Gb/sec (net) 1 GB/sec (@adapter)	10 Gb/sec (net) 6 GB/sec (@adapter)	1 Tb/sec (net) 30 GB/s (@adapter)
Memory Req.	2 GB / 128 GB	16 GB - 1 TB	1 TB - 64 TB

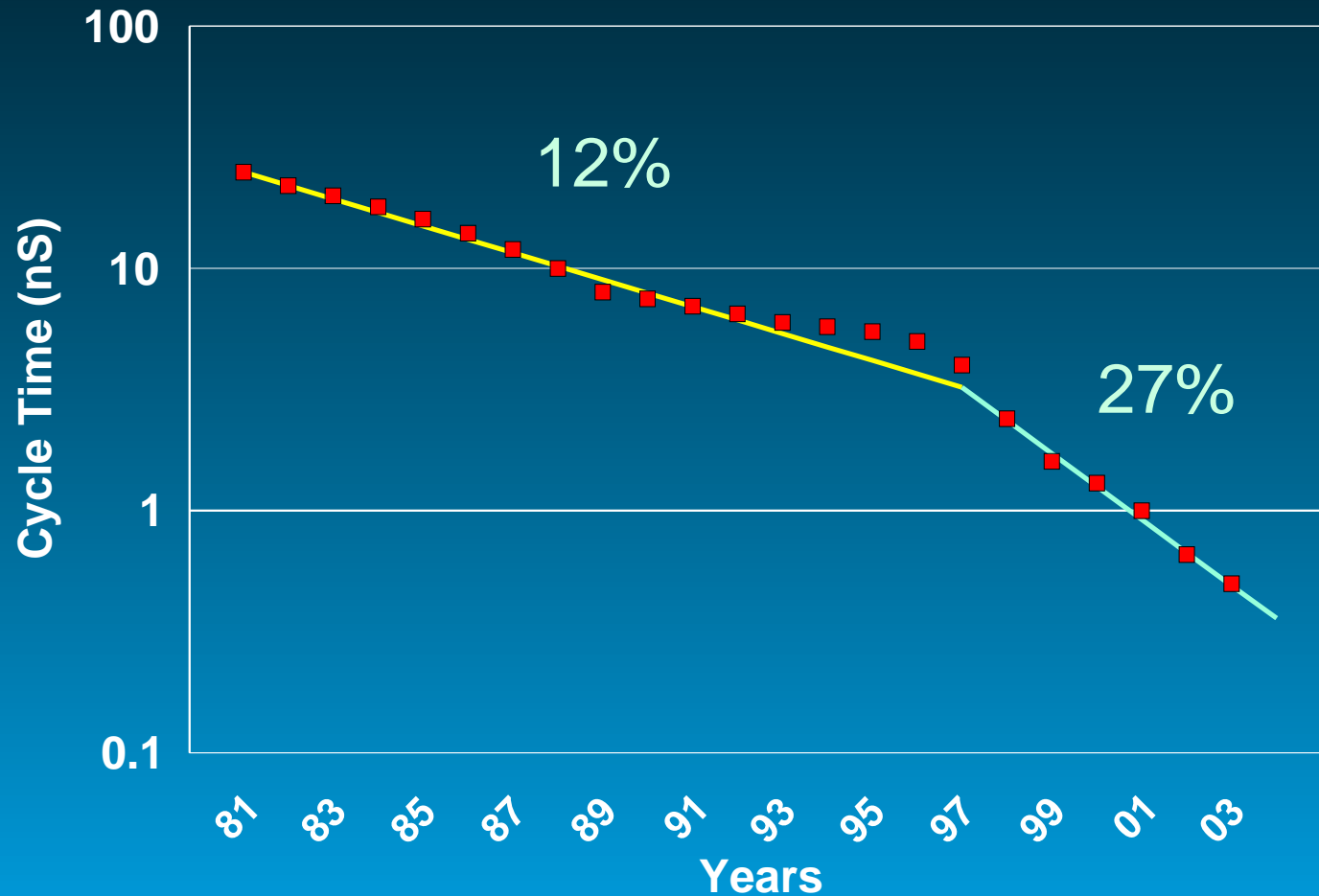
Processor Technology

Peak processor throughput growing at 60% CGR (single processor)
Peak system throughput growing at 100% CGR (increasing MP size)



Cycle Time Improvements Accelerating

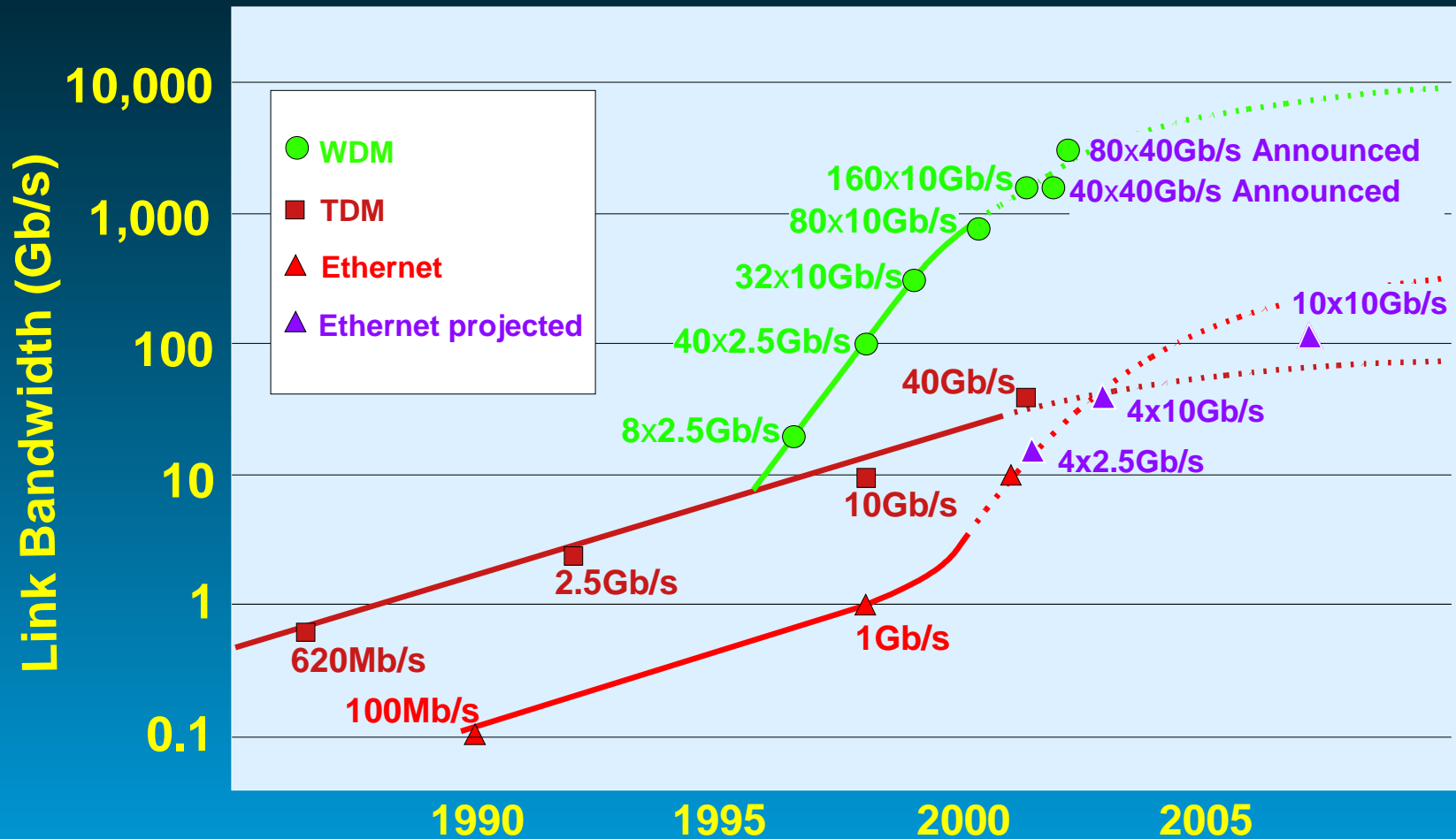
Processor Cycle Time



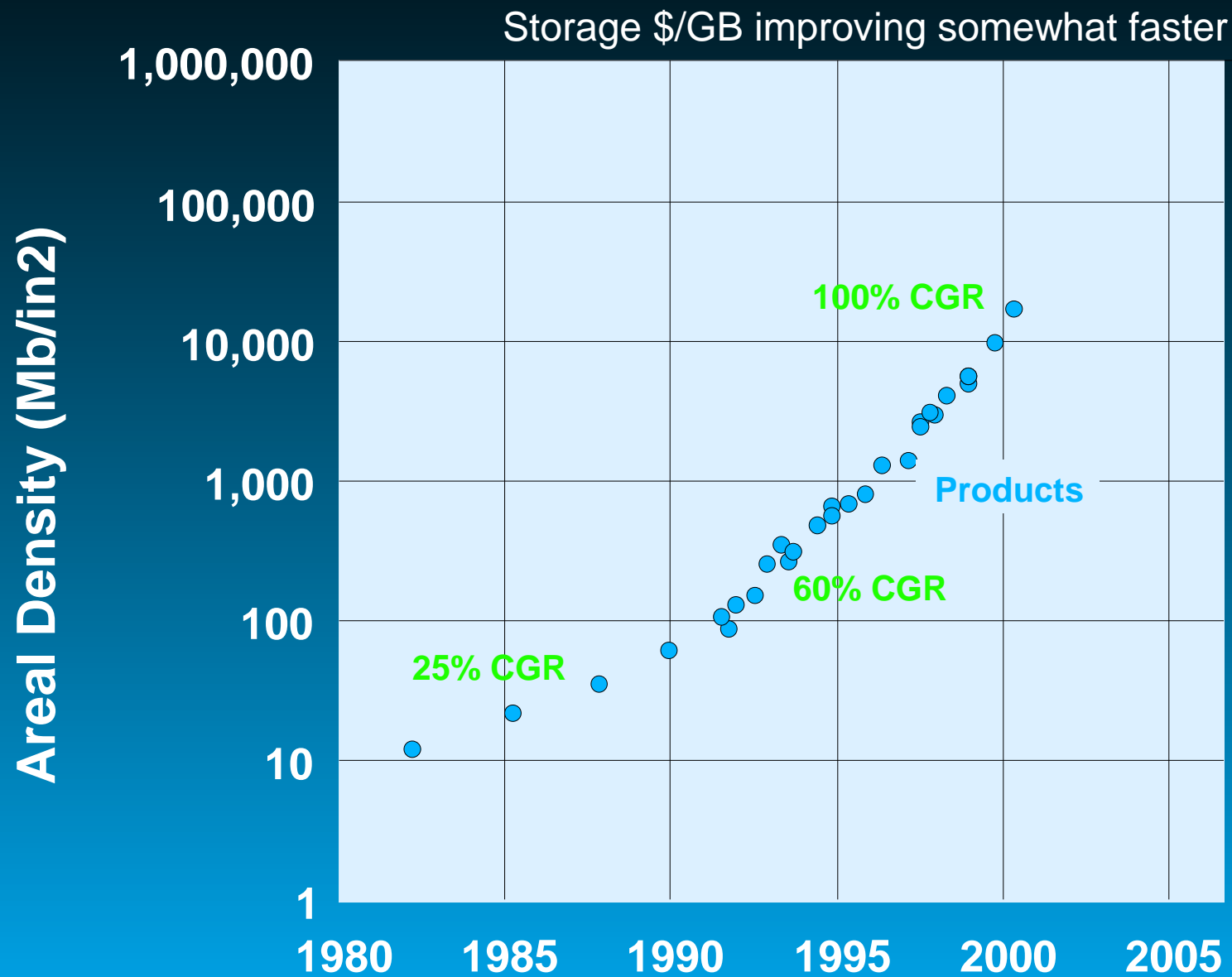
Networking Story

100,000

Networking speeds and costs improving >100%/year



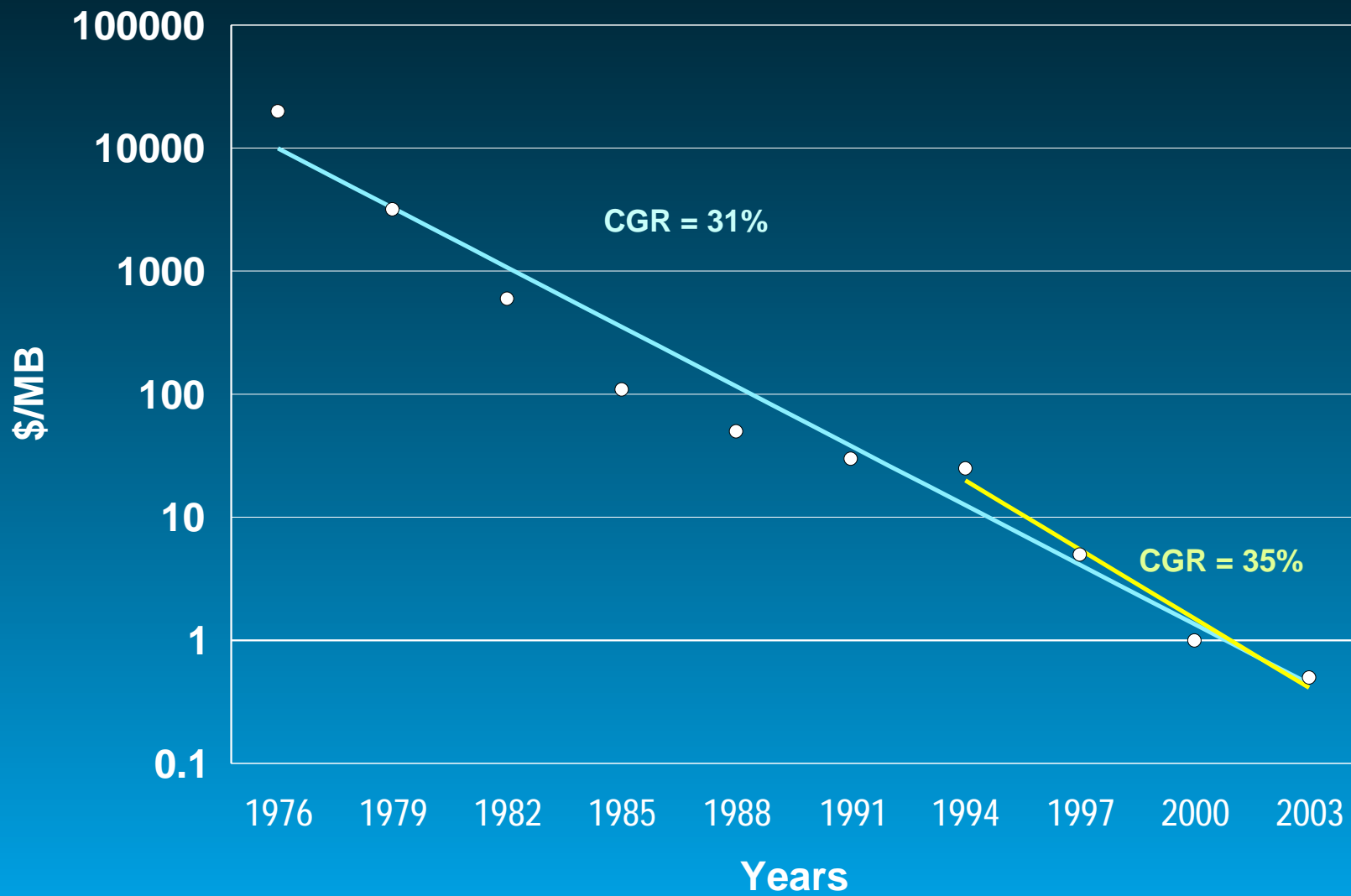
Storage Density Now Growing at 100%/year



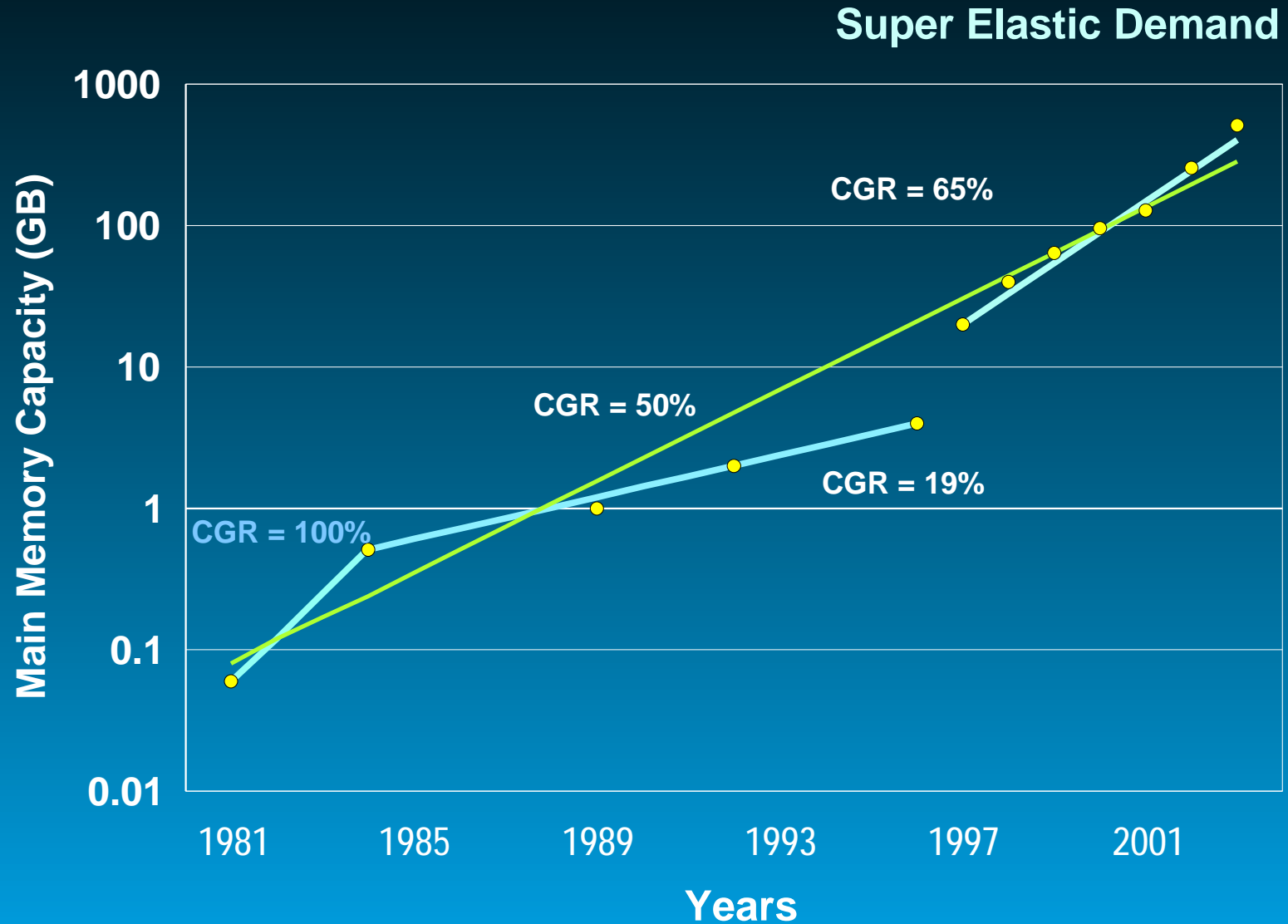
In Contrast DRAM Technology

- **Memory costs are improving slowly**
- **Memory latency is improving even more slowly**
- **Fundamental fabrication limits seem closer**
- **Increasing quantities are being used to compensate for latency to disk (also improving very slowly)**

DRAM Price Trends



Memory Sizes Growing Rapidly



Cost of memory growing!

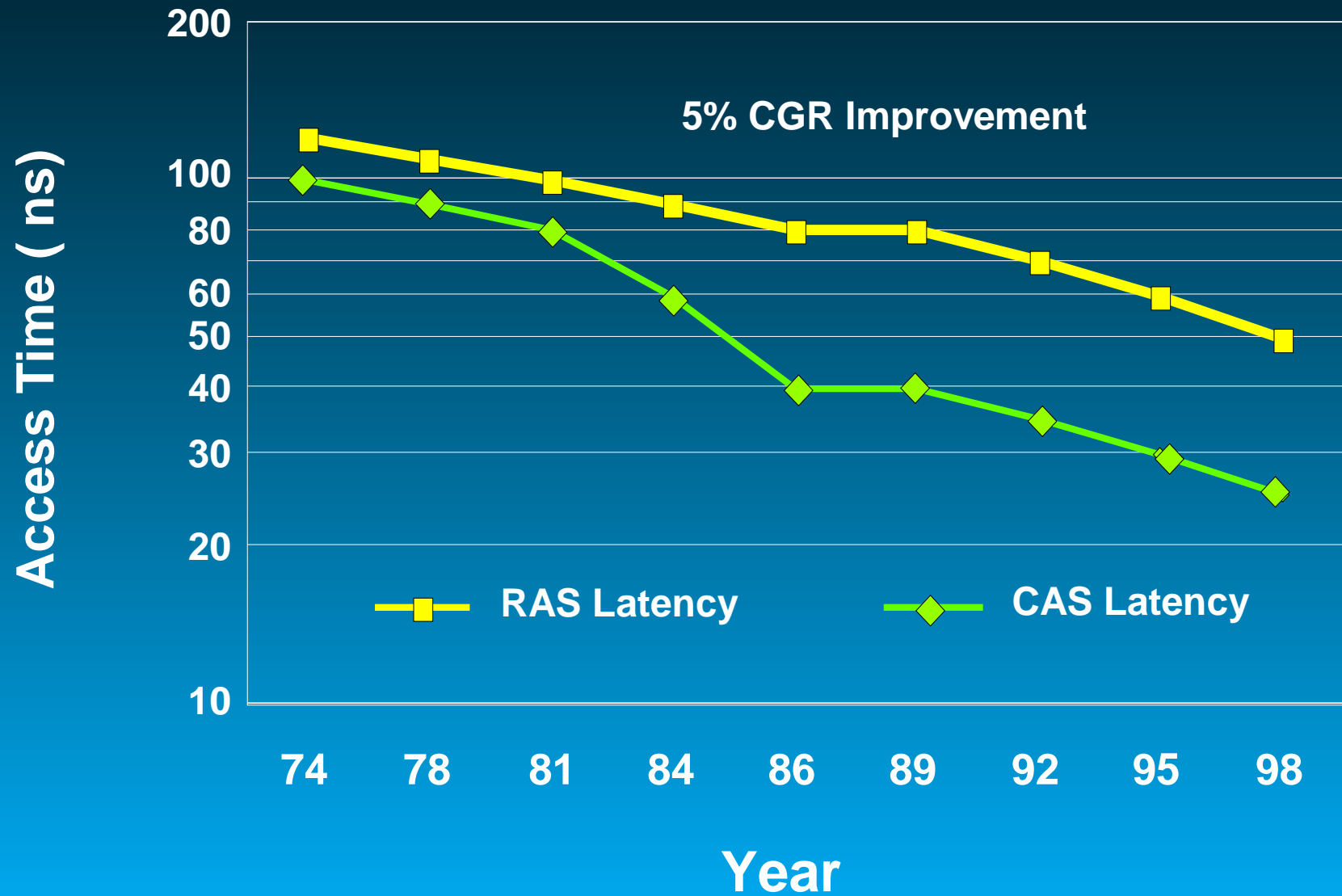
Cost of memory in typical system

growing at about 30% / year! (long term)

Single most costly single component of system excluding storage subsystem

Single most costly single component of storage subsystem (including disks themselves)

DRAM Latency



Latency to memory growing!

Latency to memory in typical system:

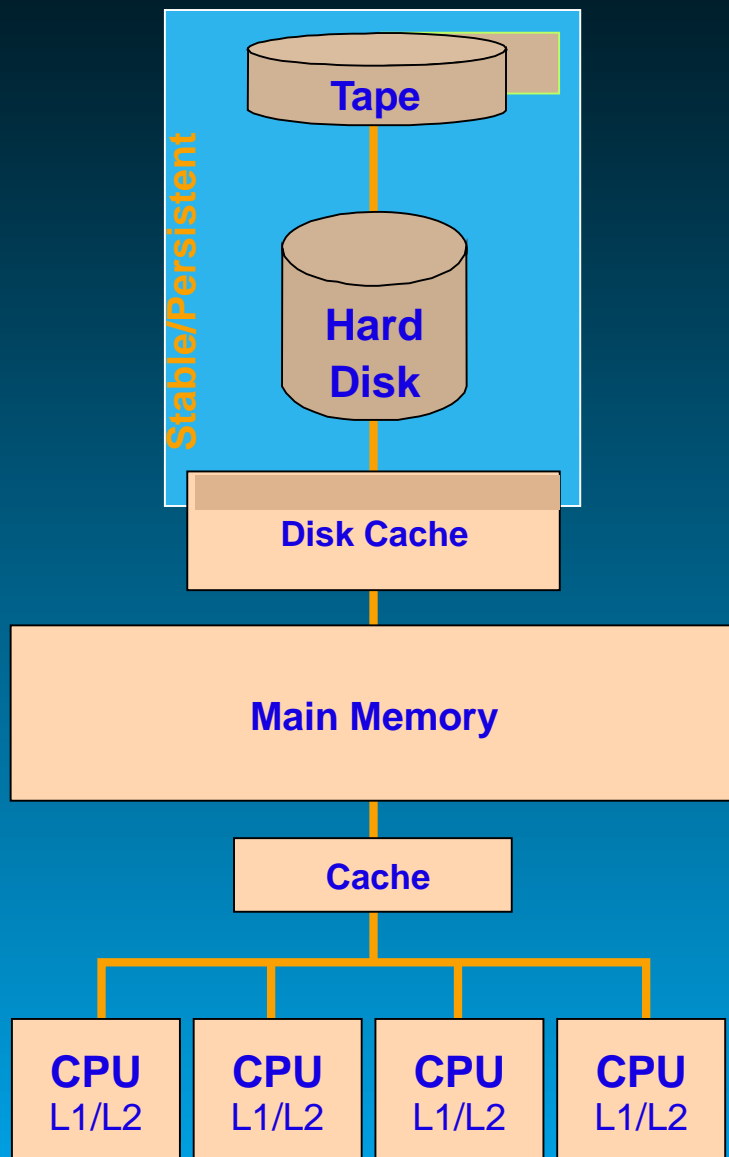
growing at about 20%-25% / year!

(relative to processor speed)

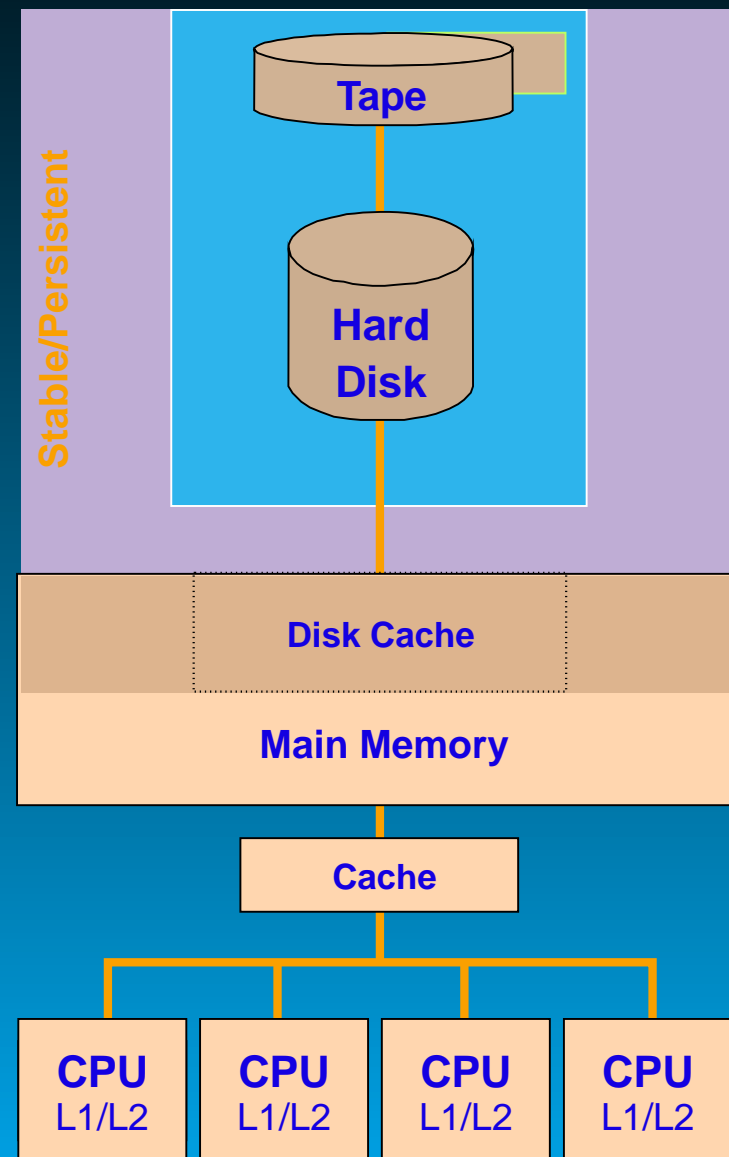
(doubles every 3-4 years)

Moving Persistent Storage Closer

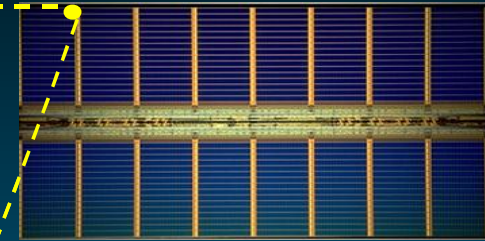
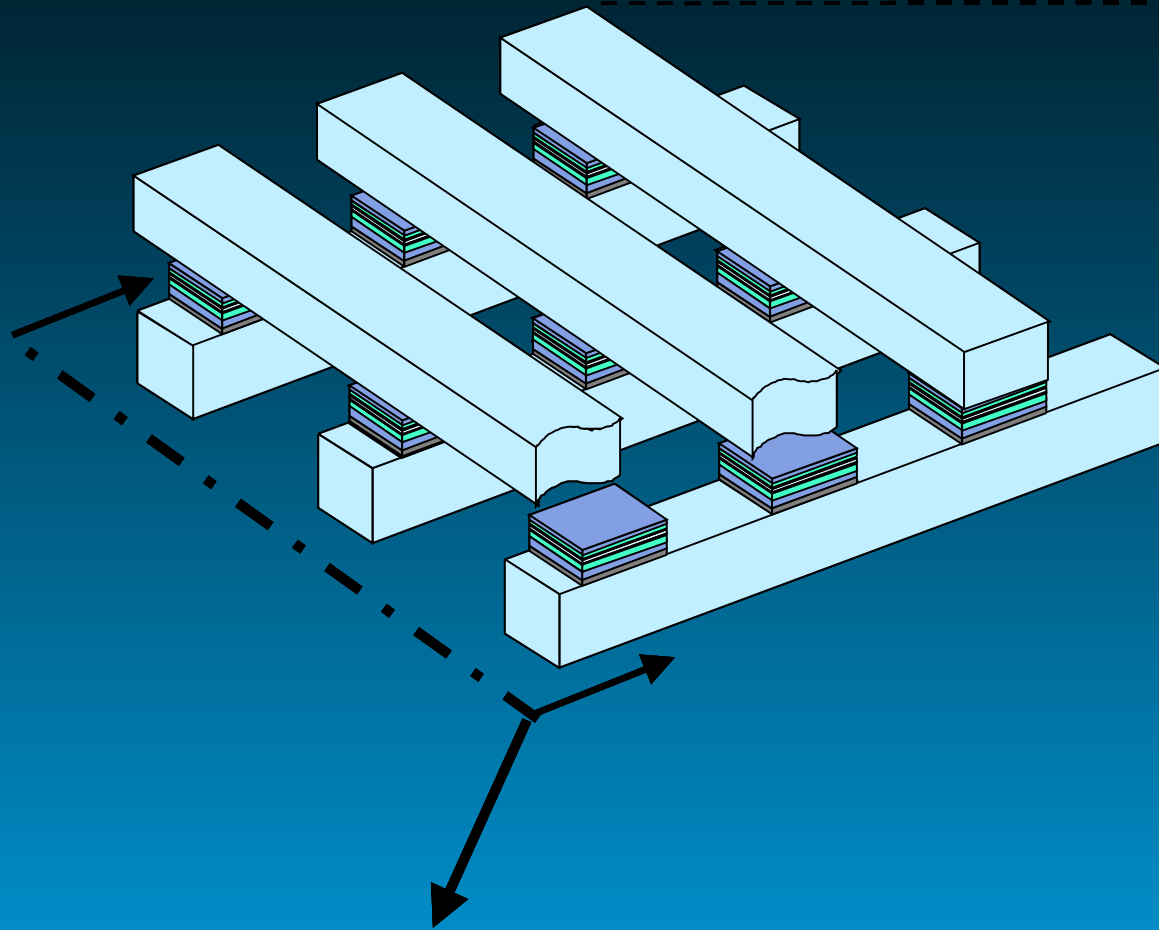
NOW



FUTURE?



MRAM: Silicon chip with magnetic memory



Cheap(er)

Permanent

Small

Fast

It's the memory!

- **Do we need a new class of memory for large caches?**
 - Larger and cheaper than SRAM caches
 - Faster and not much more expensive than DRAM
 - ✓ DRAM devices optimized for cache functions
- **Is there any hope if really deep cacheing doesn't work?**
- **Need to pull disk persistence into memory**
 - Shifting data up in the hierarchy is easiest way to improve performance
 - Will MagRAM arrive and inspire this transition?
- **Large memories imply sharing**
 - Analogy, large disk farms imply sharing
 - What is best memory organization for shared memory view of world
 - ✓ fault and failure considerations
 - ✓ consistency and locking challenge
 - ✓ efficiency, efficiency, efficiency!