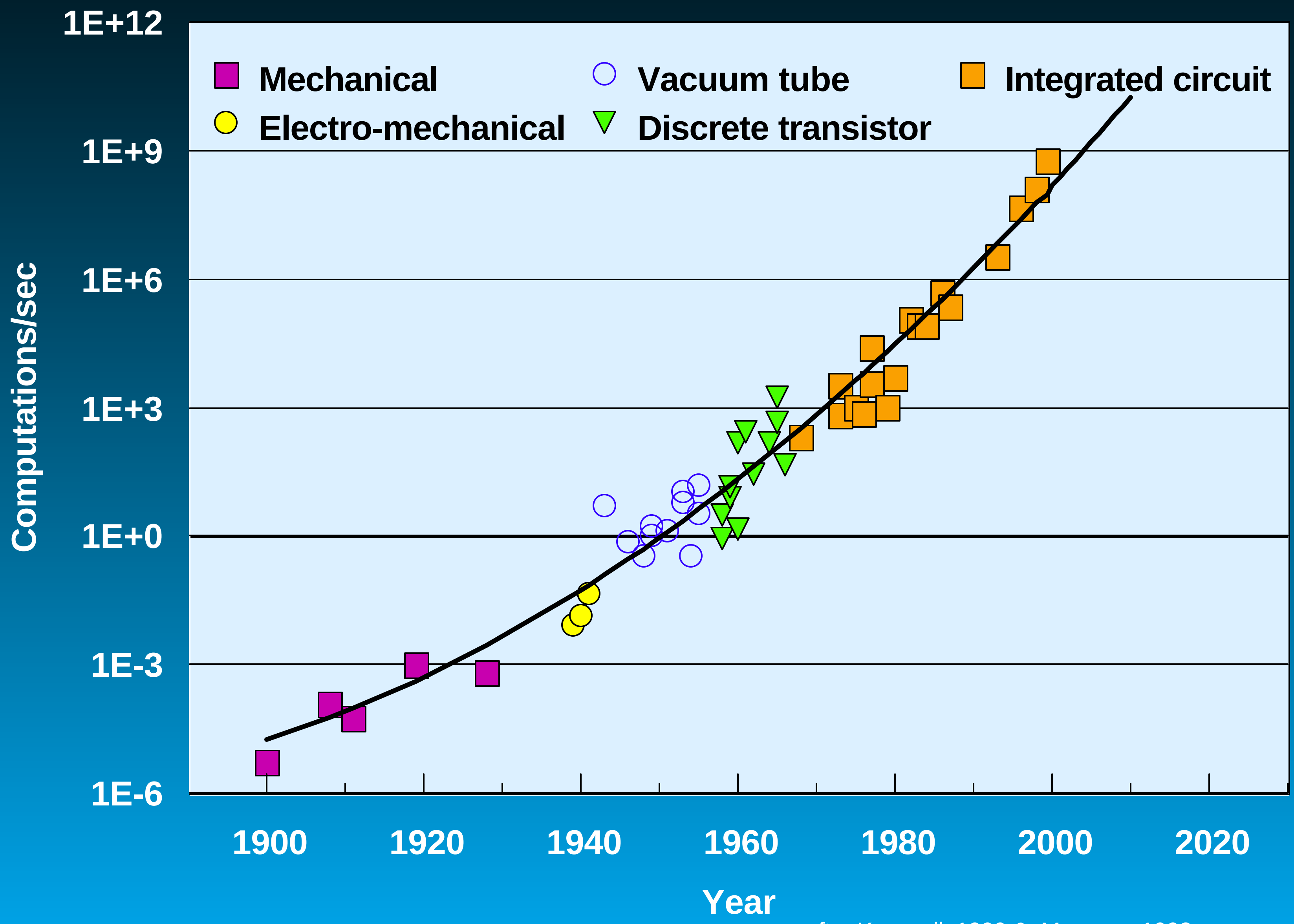
Influence of Technology Directions on System Architecture

Dr. Randy Isaac
VP of Science and Technology
IBM Research Division

September 10, 2001

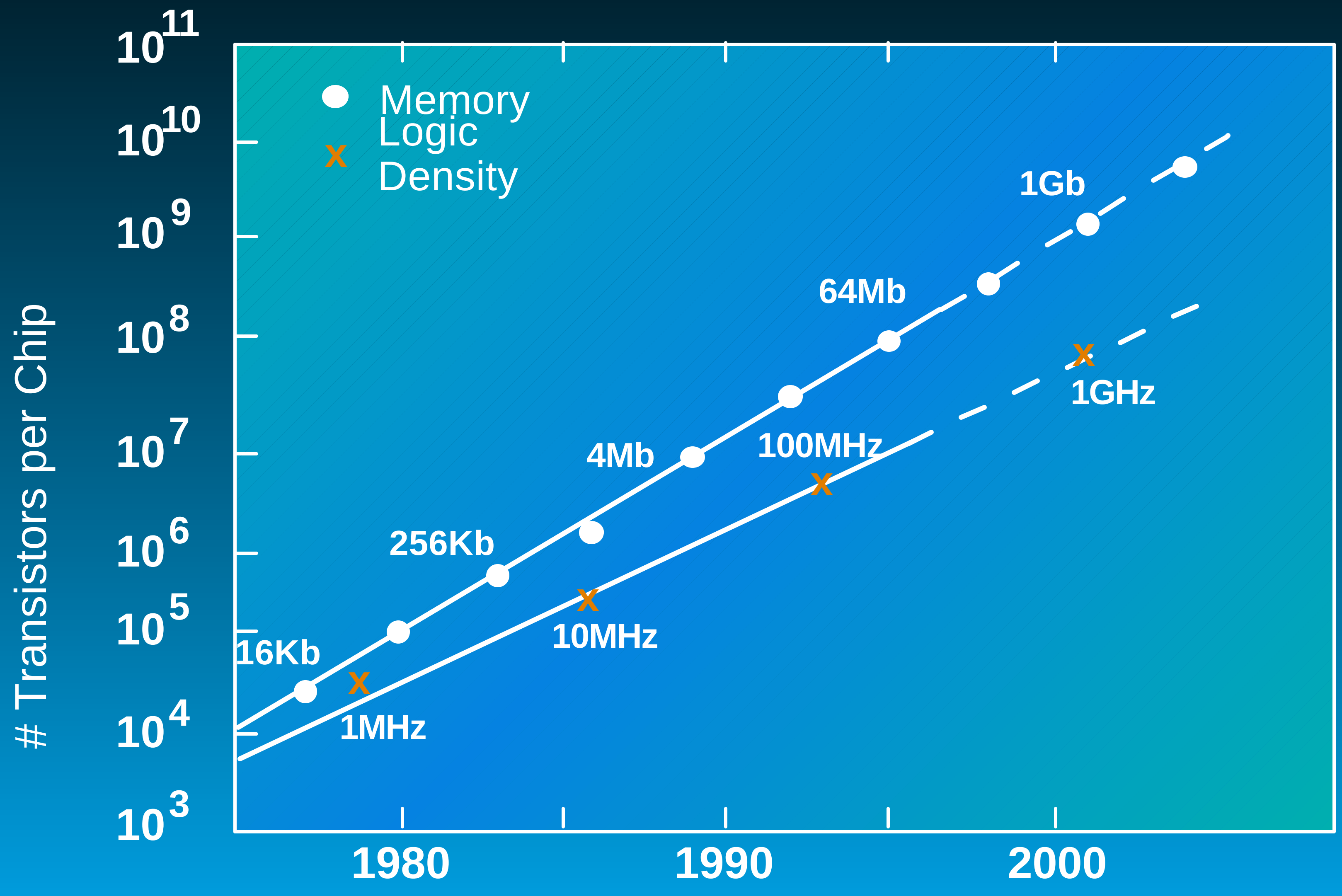
- Moore's Law continues beyond conventional scaling
- Power becomes the limiting metric
- The integration focus moves from circuit to processor

$1000 Buys...

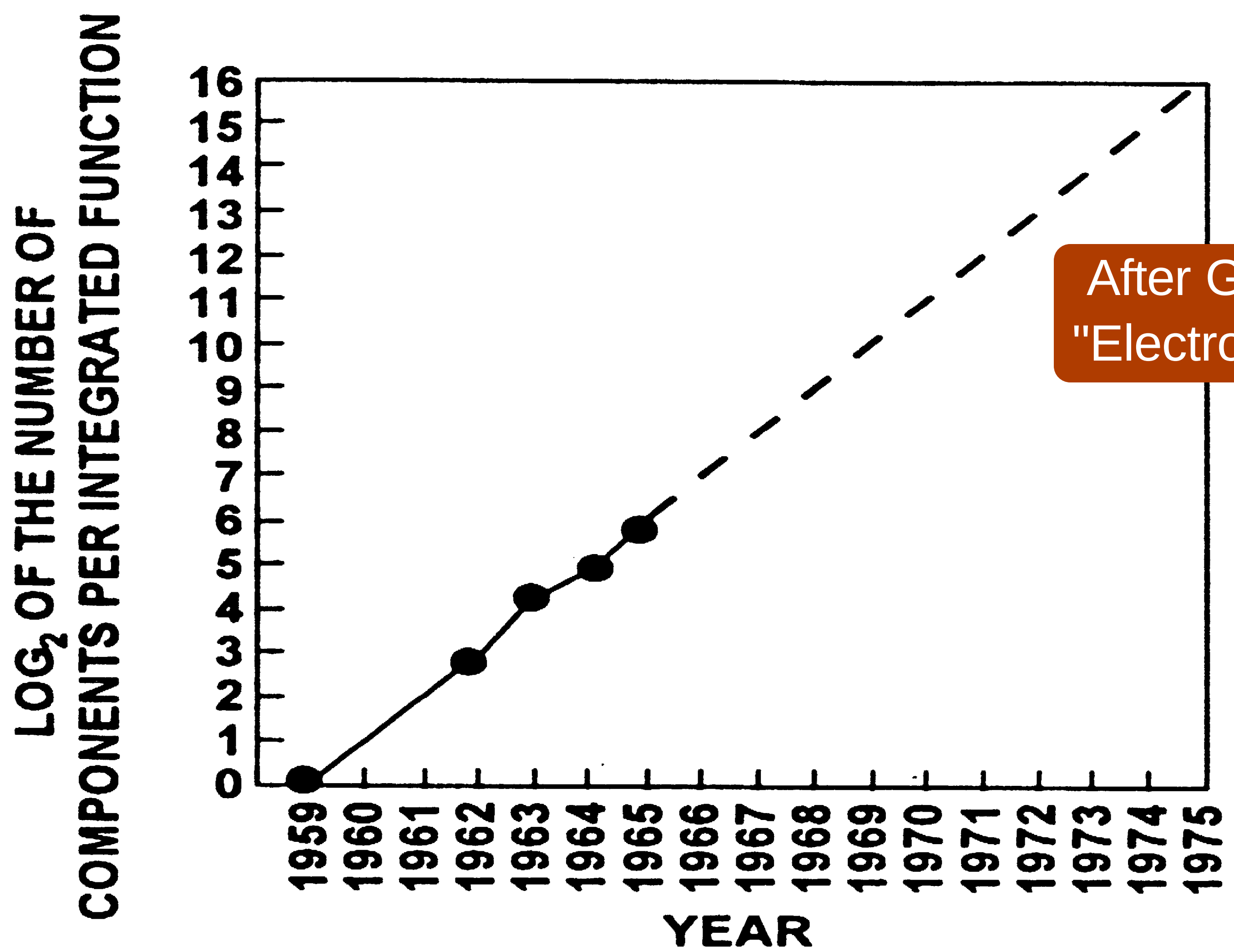


after Kurzweil, 1999 & Moravec, 1998

Integrated Circuit Performance Trends

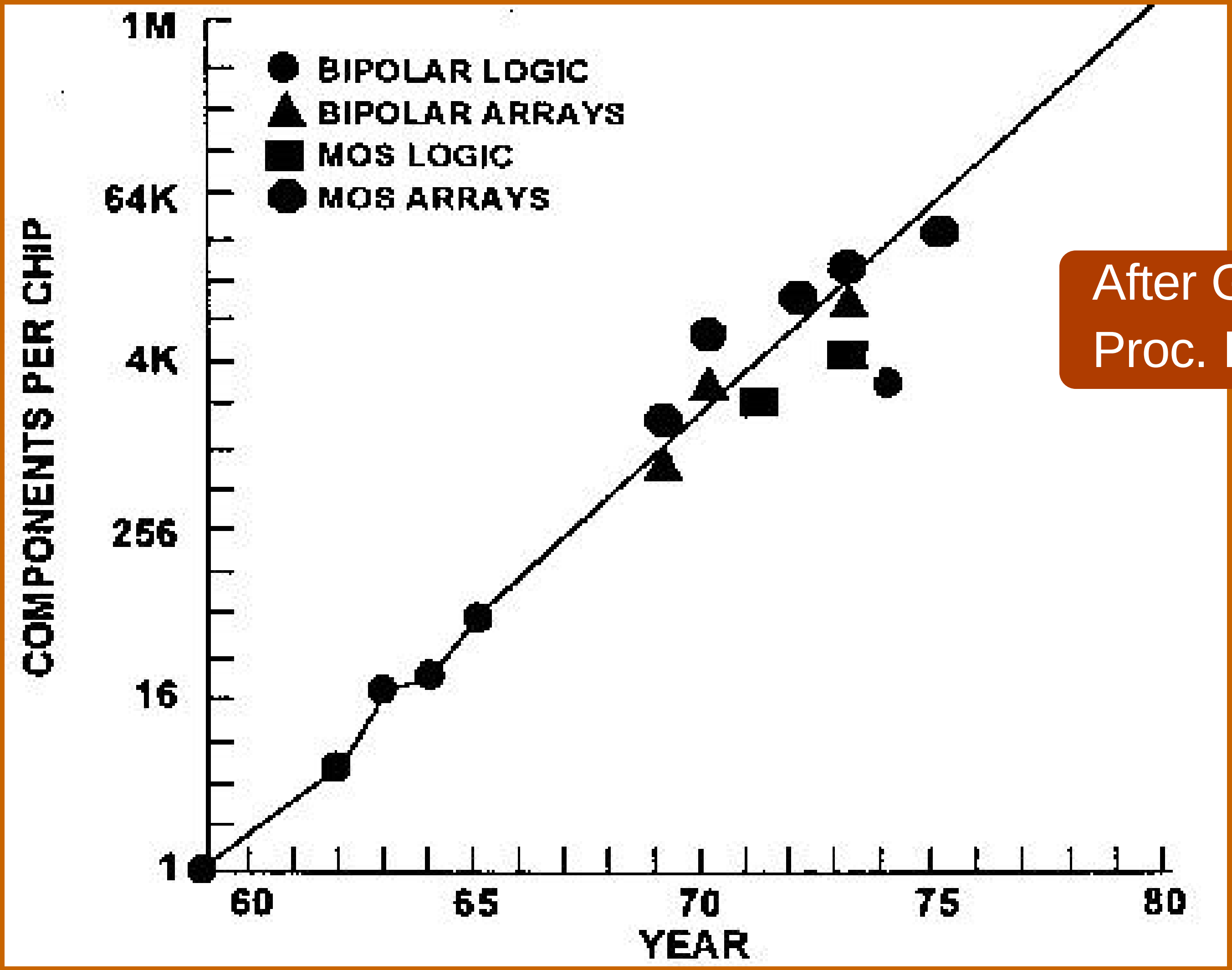


The Original Moore's Law Proposal



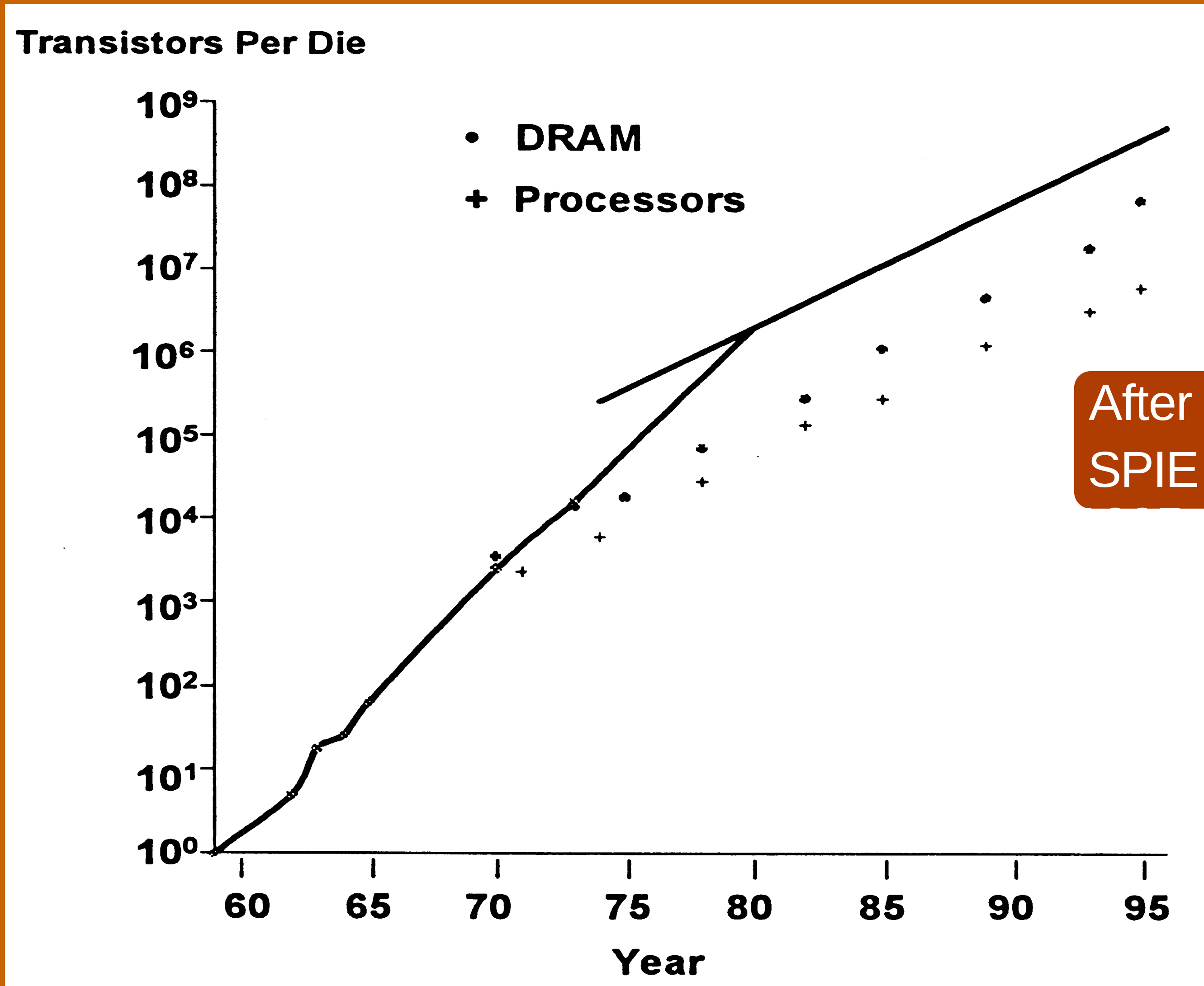
After G. E. Moore "Electronics," 1965

A Decade of Agreement



After G. E. Moore
Proc. IEDM, 1975

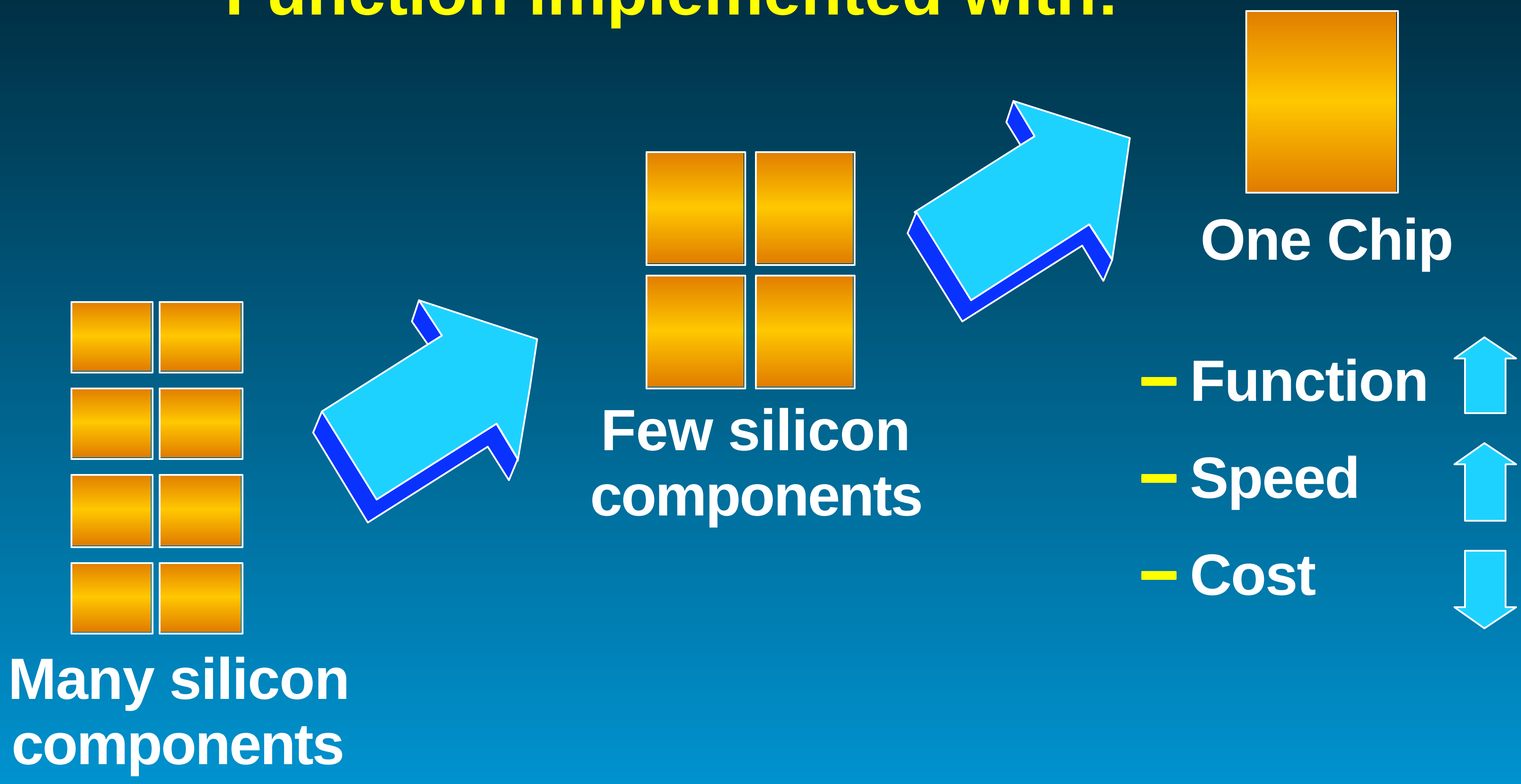
Complexity's Influence



After G. E. Moore
SPIE *v.* 2440,

Increased integration

Function implemented with:

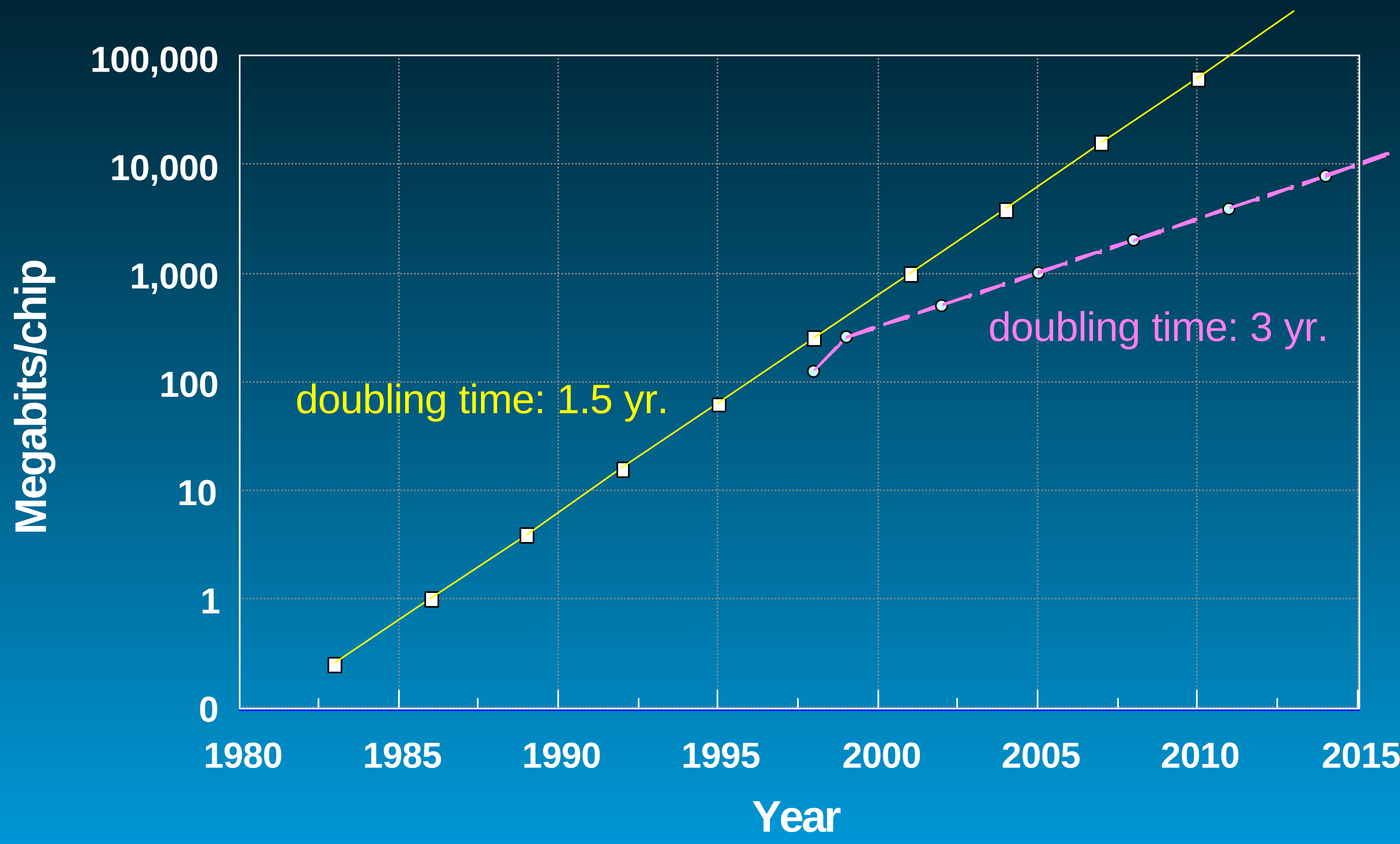


Partitioning the Improvement Rate

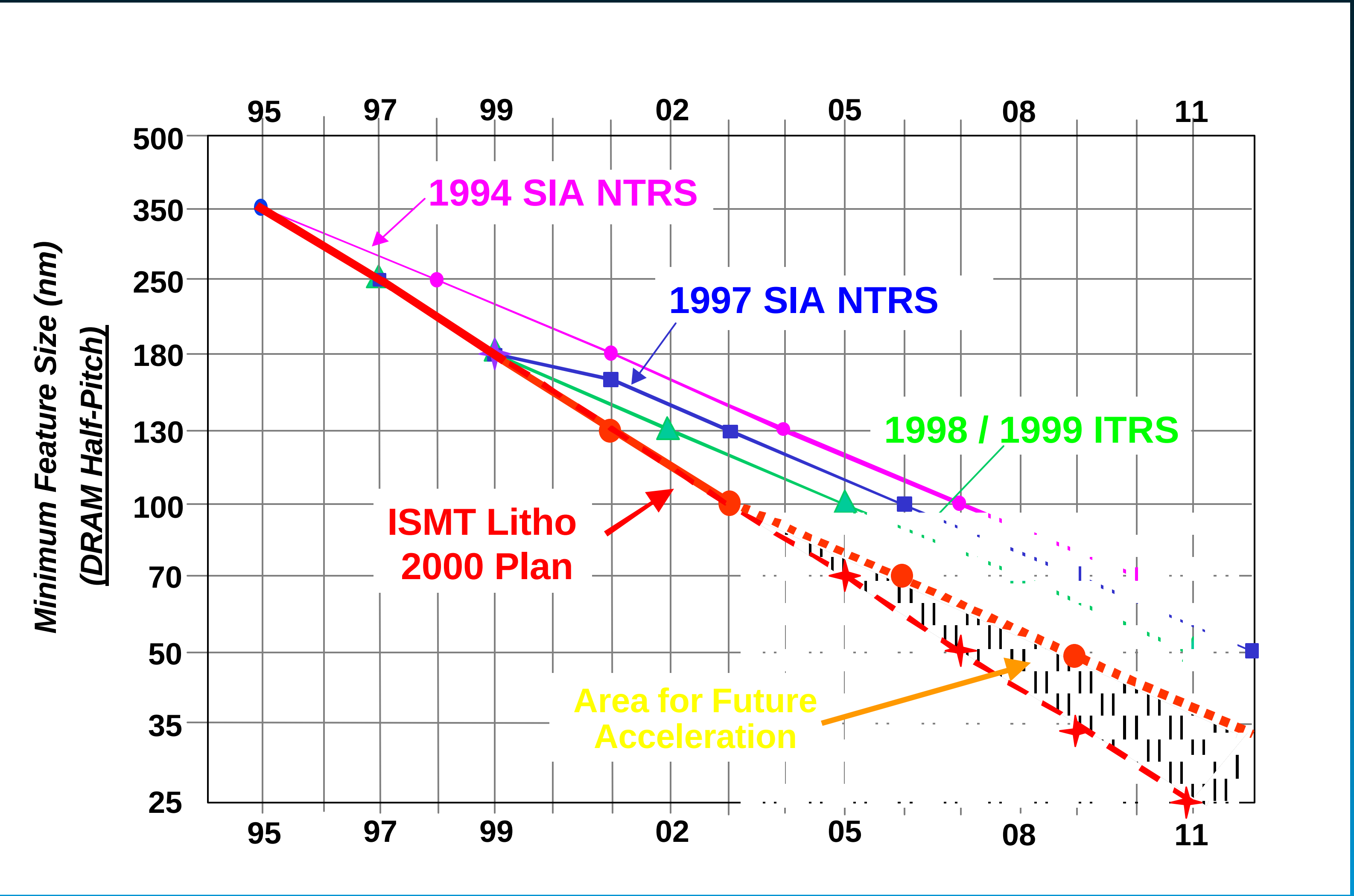
- **Improving Integration: Components per chip**
  - **50% Gain from Lithography**
  - **25% Gain from Device and Circuit Innovation**
  - **25% Gain from Increased Chip Size (manufacturability)**

- **Improving Performance:**
  - **Transistor Performance Improvement**
  - **Interconnect Density and Delay**
  - **Packaging and Cooling**
  - **Circuit-level and System-level Gains**

Evolution of Memory Density

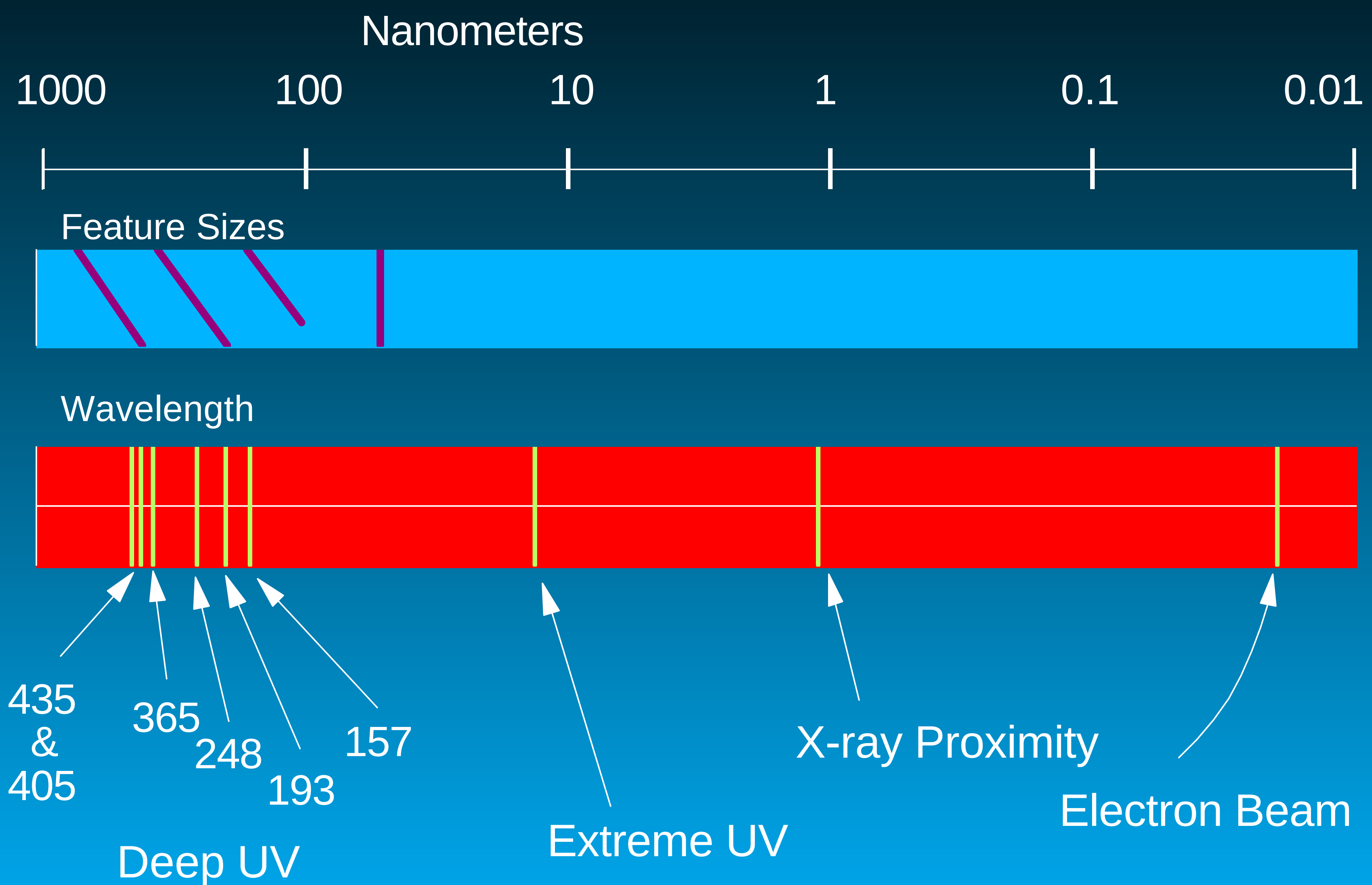


ITRS Lithography Roadmap



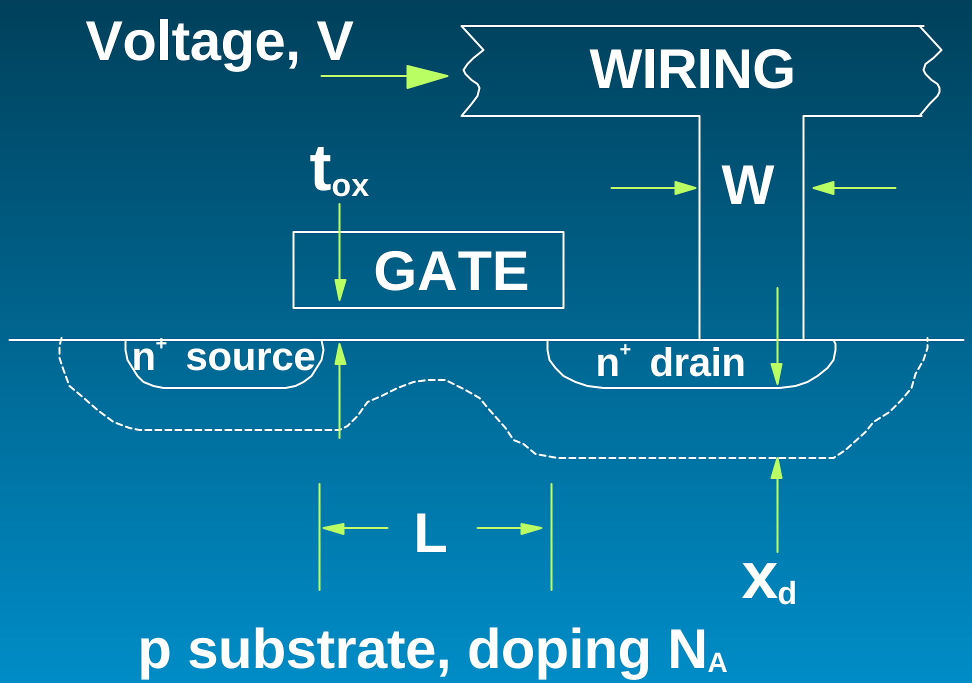
- Industry-Wide Lithography Technology Acceleration

Dimensions in Lithography



Device Scaling

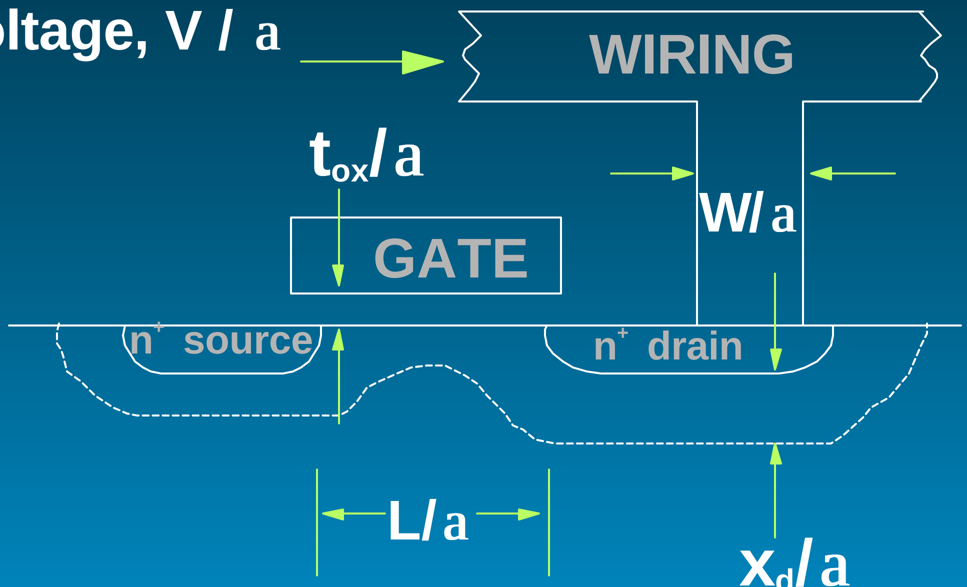
Original Device



Device Scaling

Scaled Device

Voltage, V / a



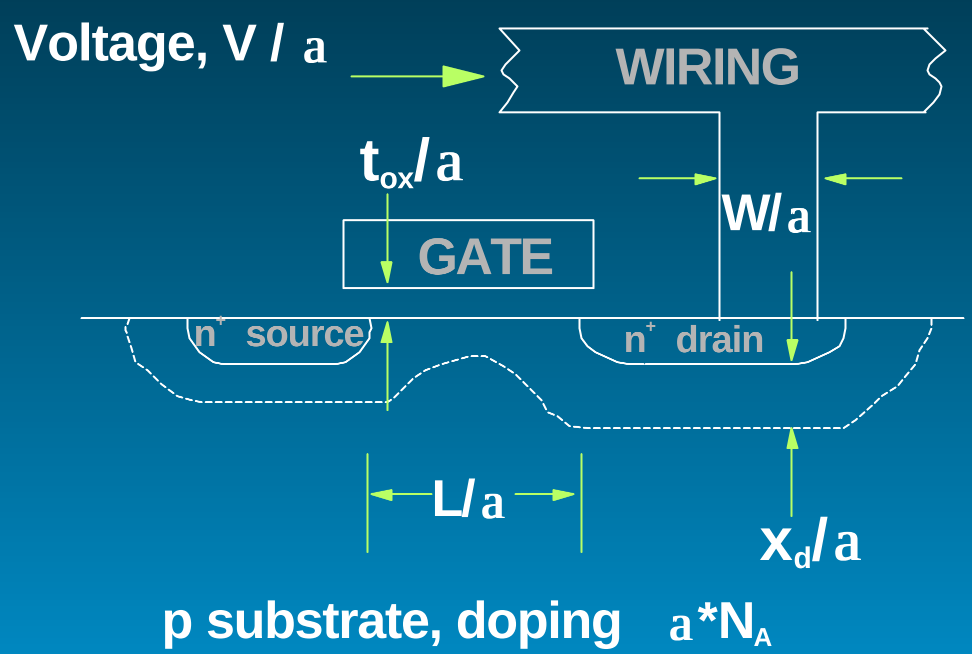
p substrate, doping a*N_A

SCALING:

Voltage: V/ a
Oxide: t_{ox} / a
Wire width: W/ a
Gate width: L/ a
Diffusion: x_d / a
Substrate: a * N_A

Device Scaling

Scaled Device



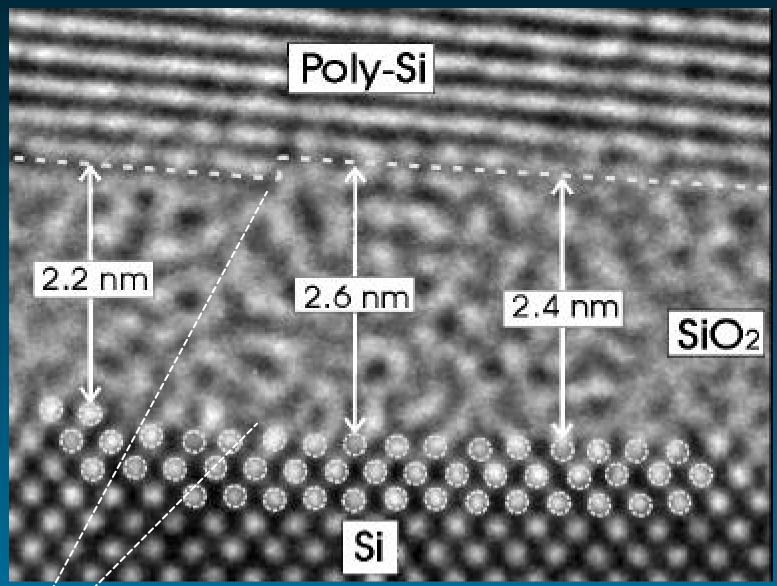
SCALING:

Voltage:	V/ a
Oxide:	t_{ox} / a
Wire width:	W/ a
Gate width:	L/ a
Diffusion:	x_d / a
Substrate:	a * N_A

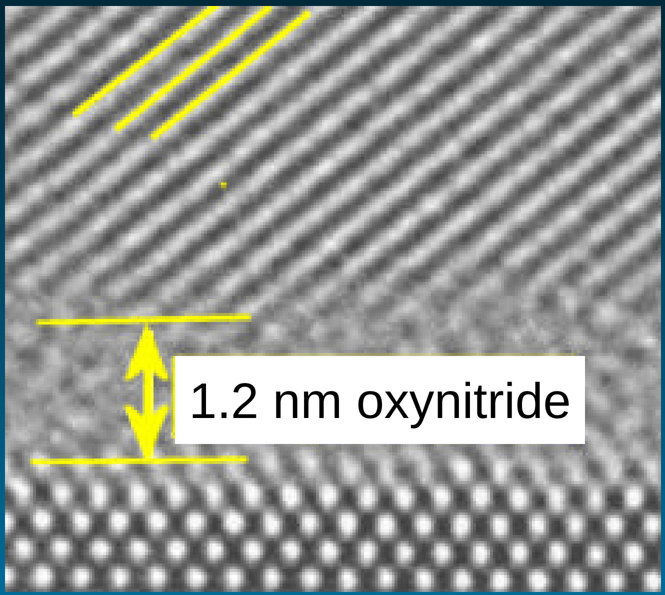
RESULTS:

Higher Density:	~ a^2
Higher Speed:	~ a
Lower Power/ckt:	~1/ a^2
Power Density:	~Constant

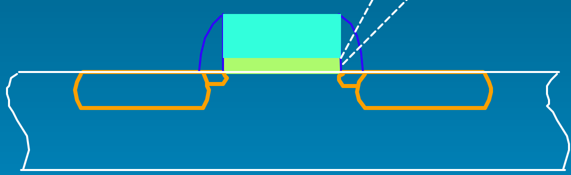
Fundamental atomic limit to scaling recipe



present



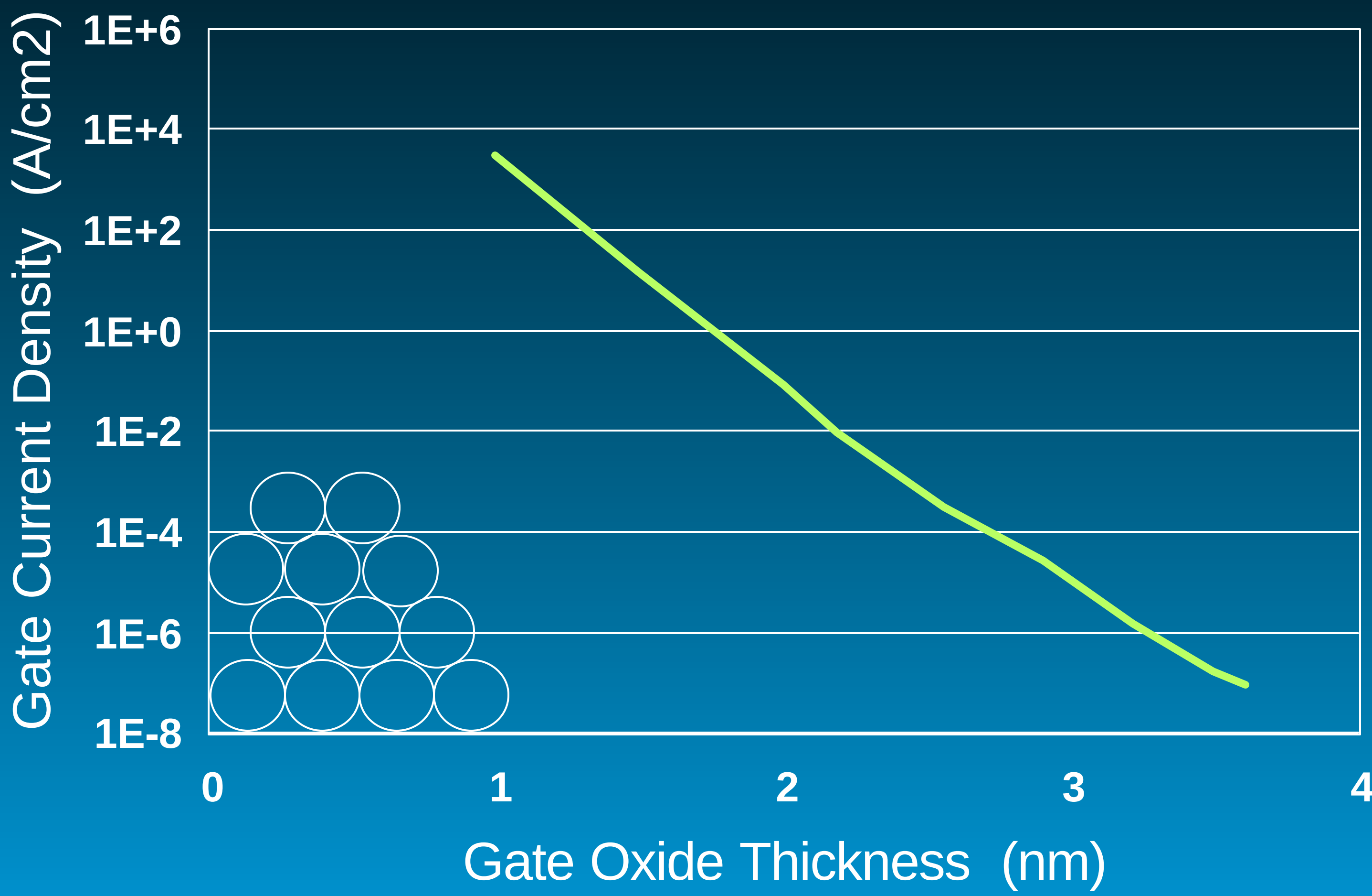
future



silicon bulk field effect transistor (FET)

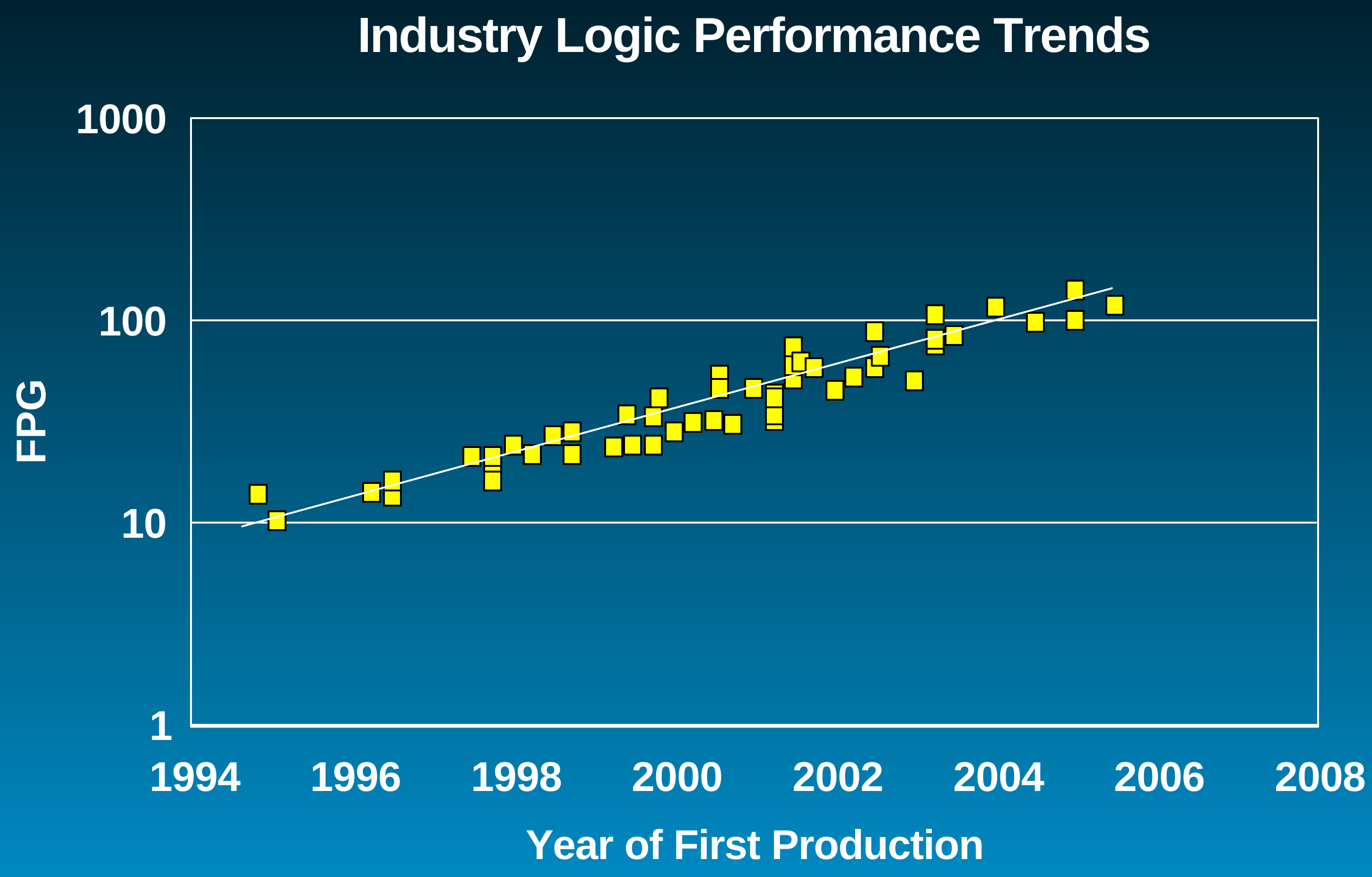
- **Oxide thickness is approaching a few atomic layers**

Limit of Oxide Scaling



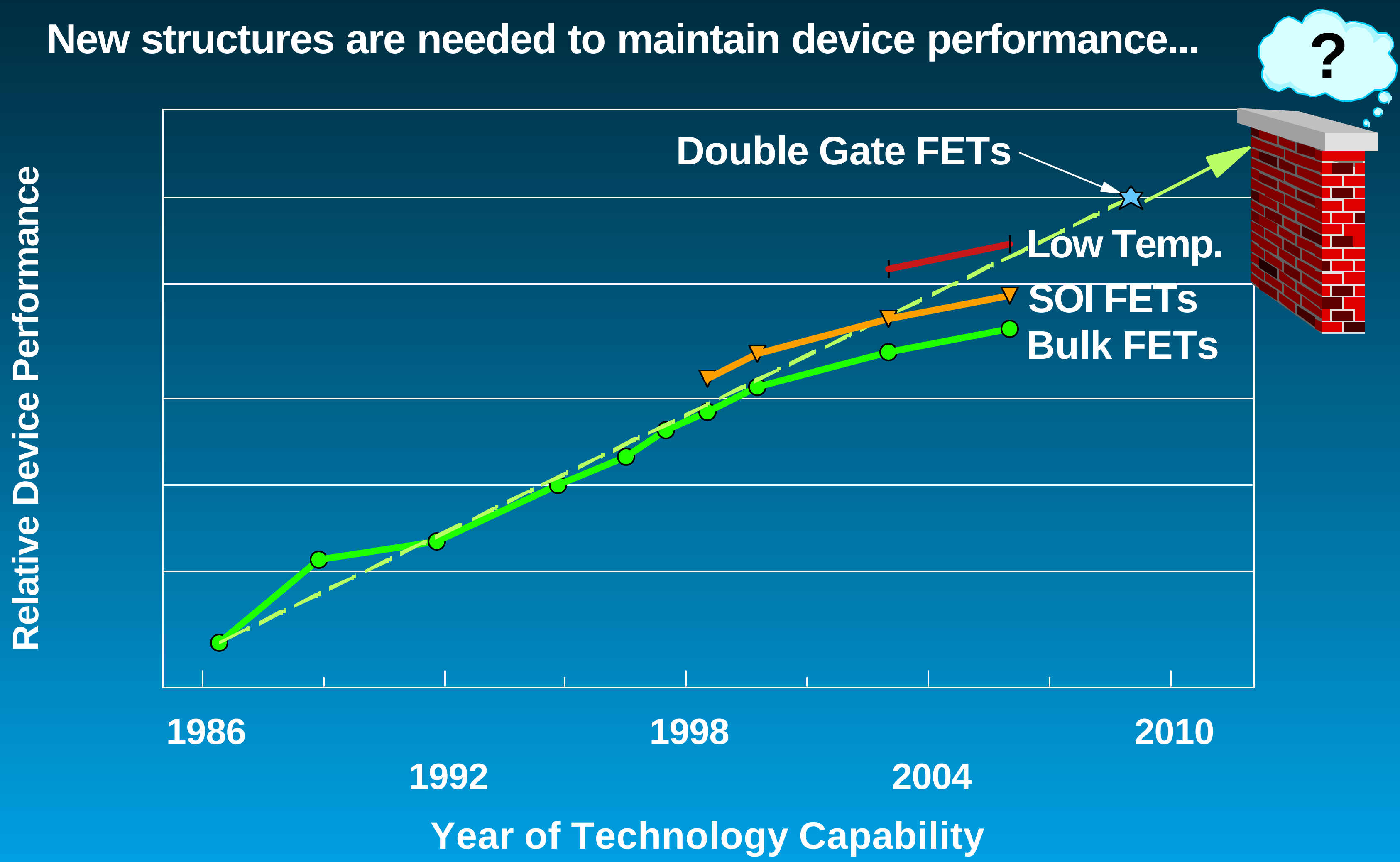
(Gate voltages: 0.9 to 2.0 V)

High Performance CMOS Logic Trend

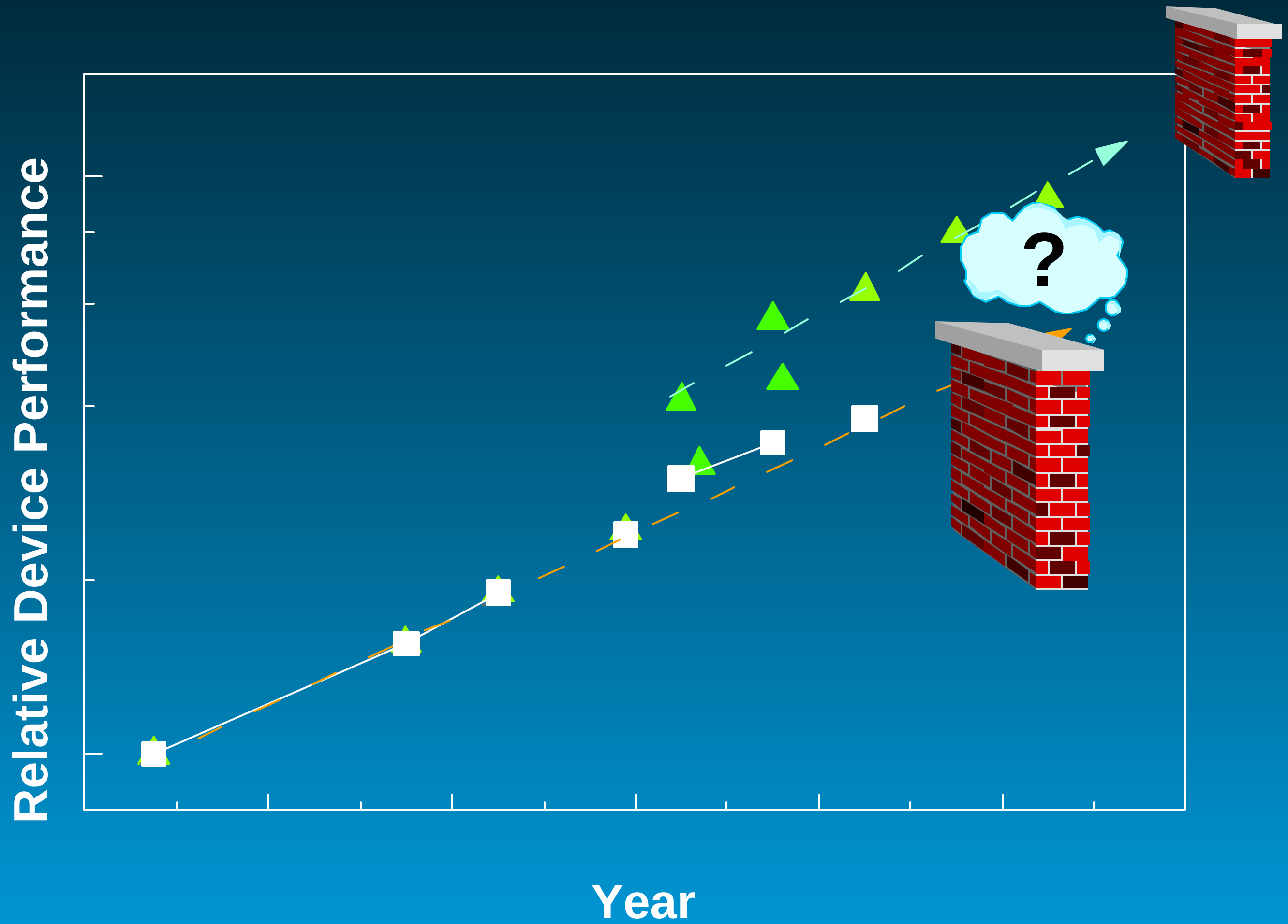


Relative CMOS Device Performance

New structures are needed to maintain device performance...

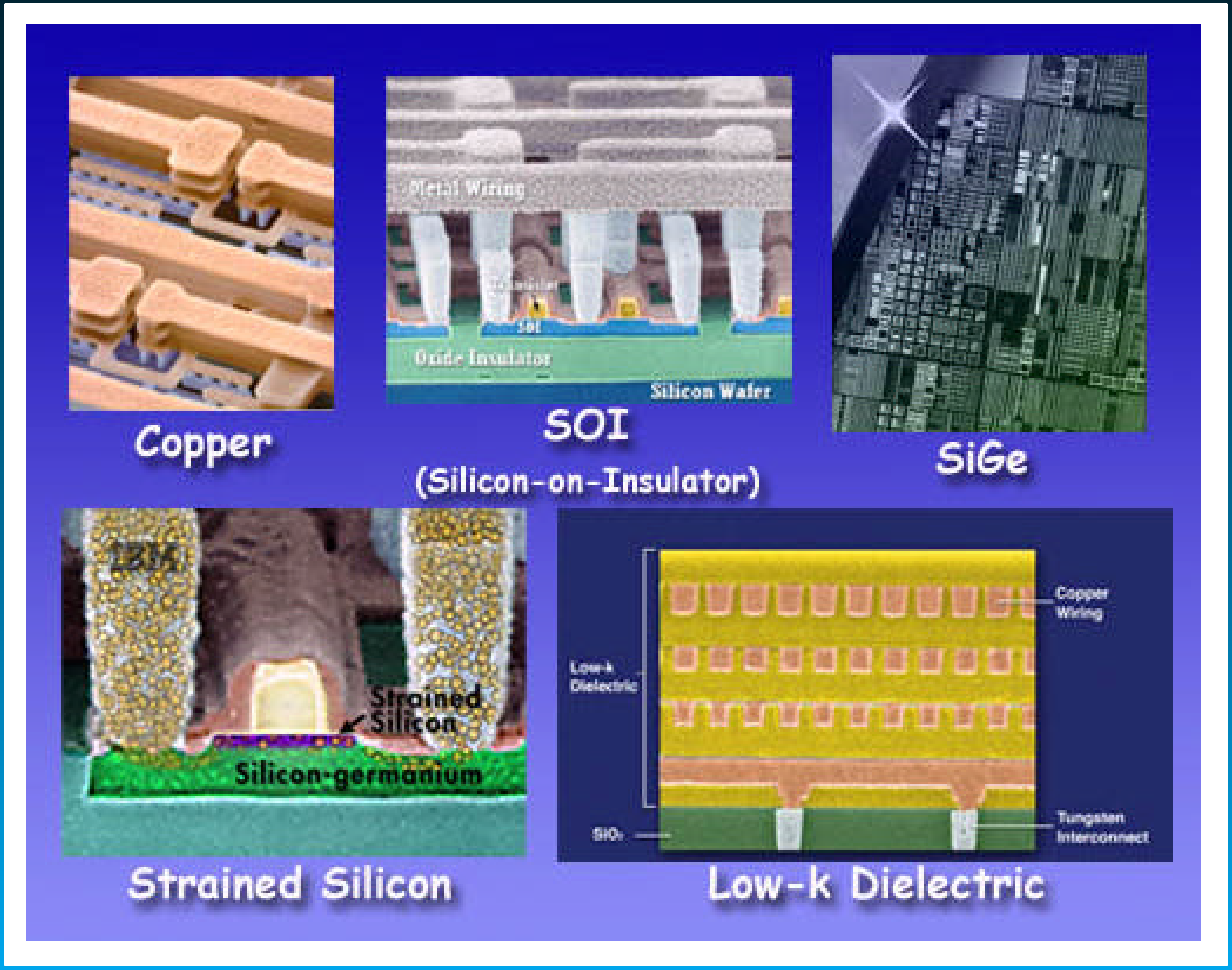


MOSFET Device Structure (R)evolution

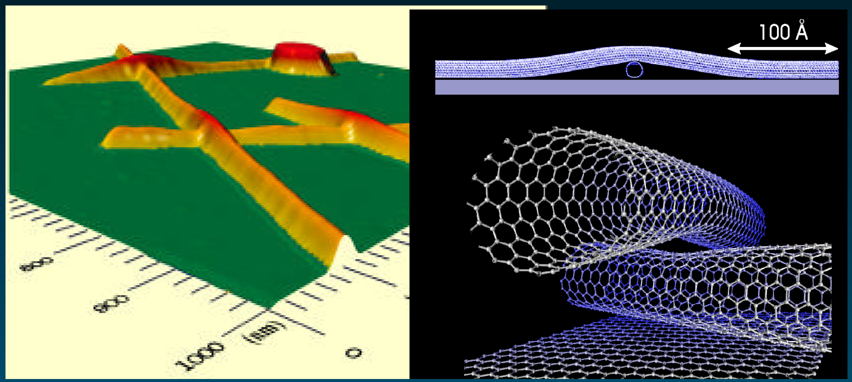


- New devices/materials support accelerated growth rate

Better Performance Without Scaling

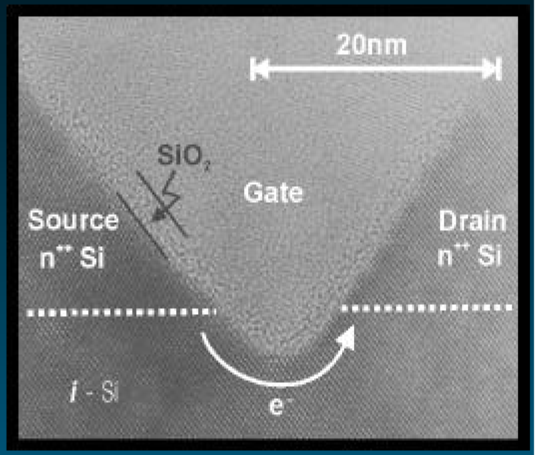


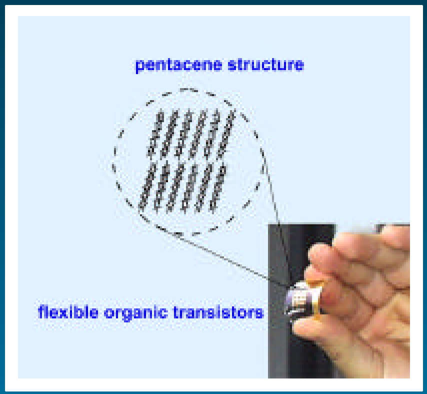
Novel Devices



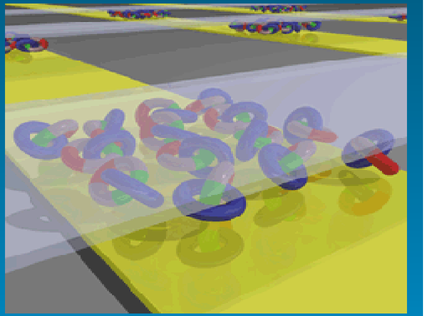
Carbon Nanotubes

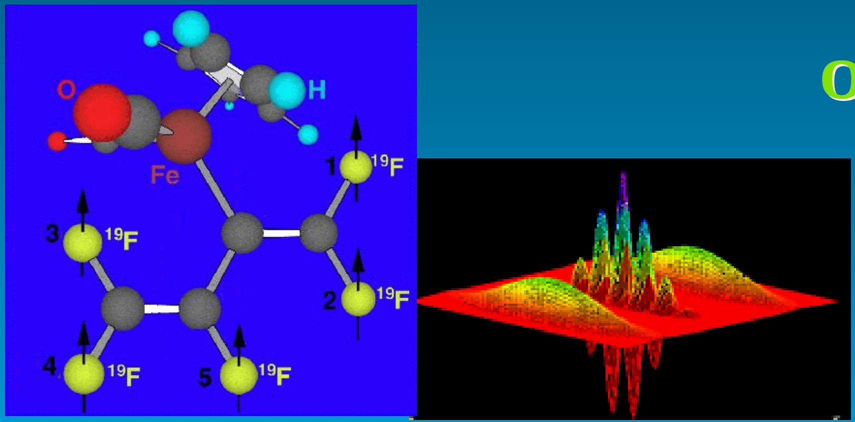
V-Groove Transistors



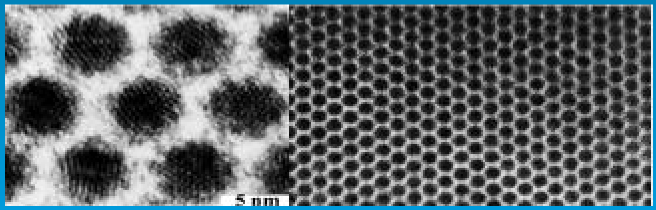


Organic Transistors





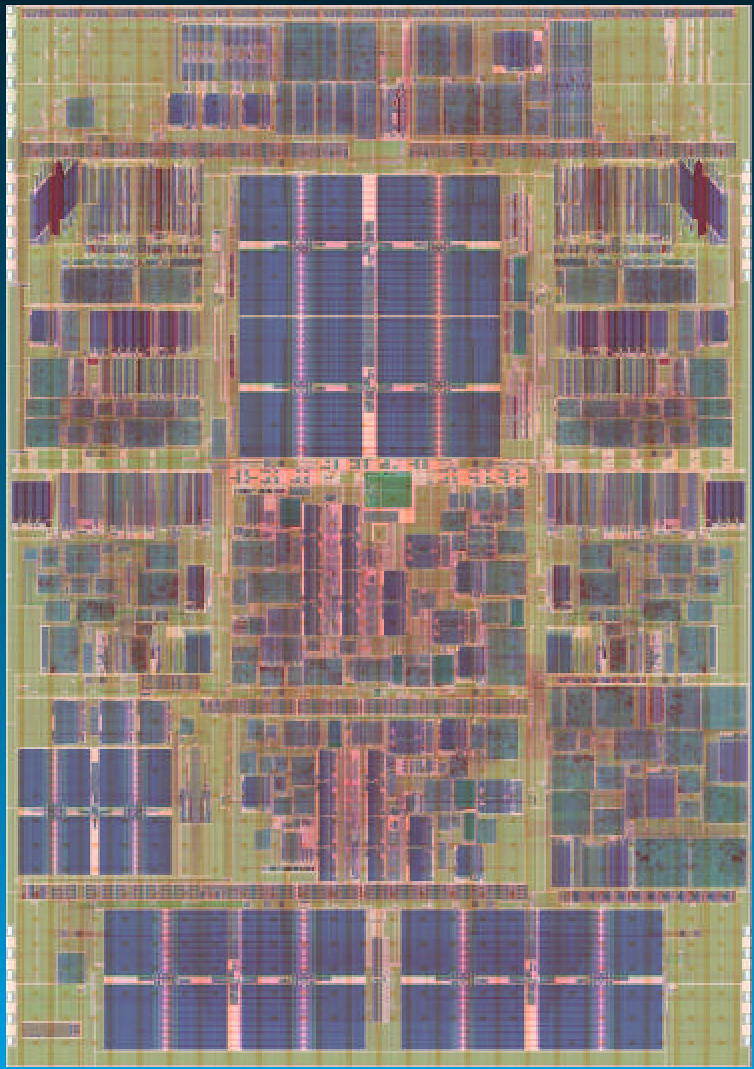
Quantum Computing



Molecular Devices

64-bit S/390 Microprocessor

- 47 Million transistors
- Copper interconnect -- 7 layers
- Size: 17.9 x 9.9 mm
- Single scalar, in-order execution
- Split L1 cache (256K I & D)
- BTB 2K x 4, multiported
- On chip compression unit
- > 1 GHz frequency on a 20-way system

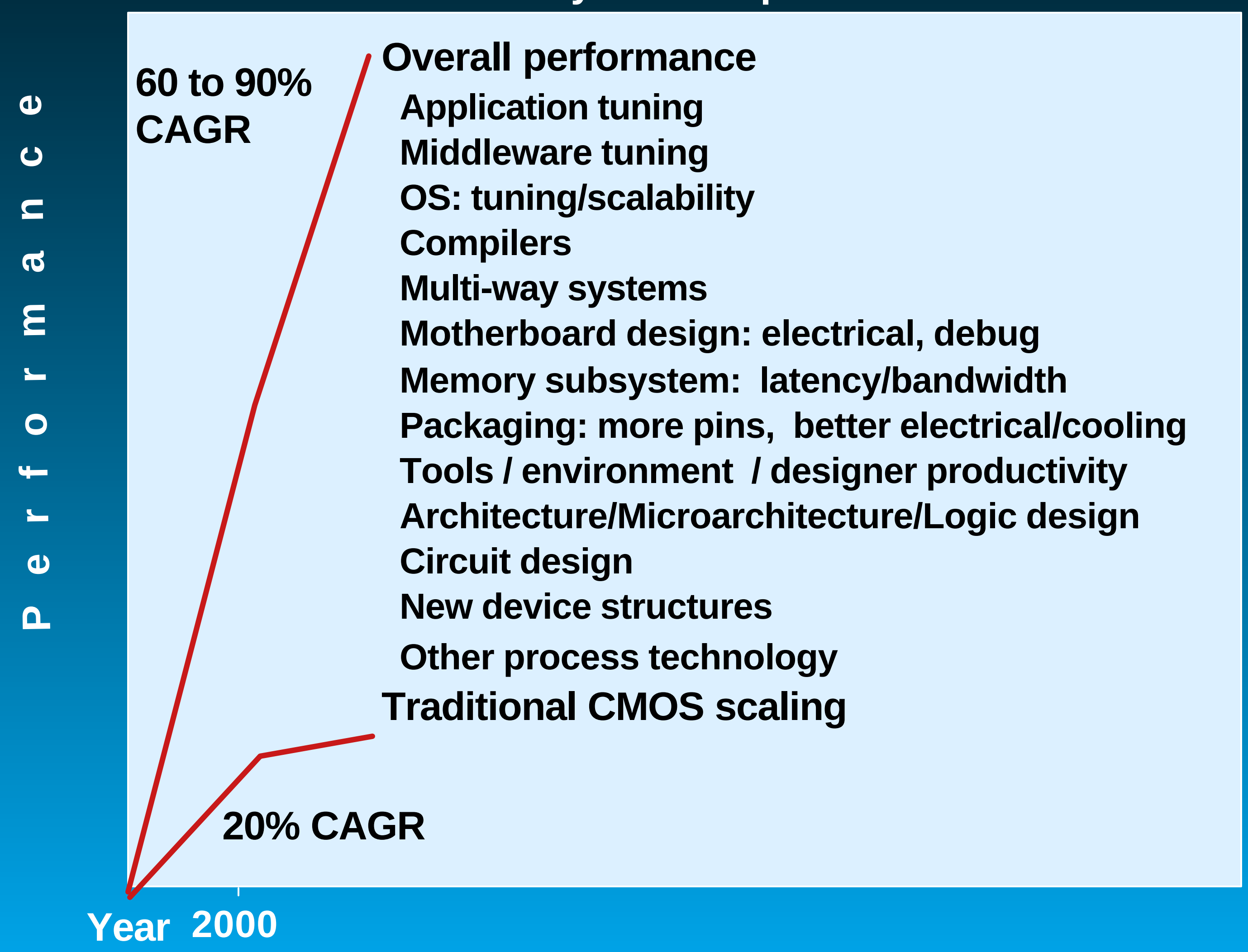


Blue Pacific

- 3.9 trillion operations/sec
- Can simulate nuclear devices
- 15,000 X speed of average desktop
- 80,000 X memory of average desktop
- 75 terabytes of disk storage capacity

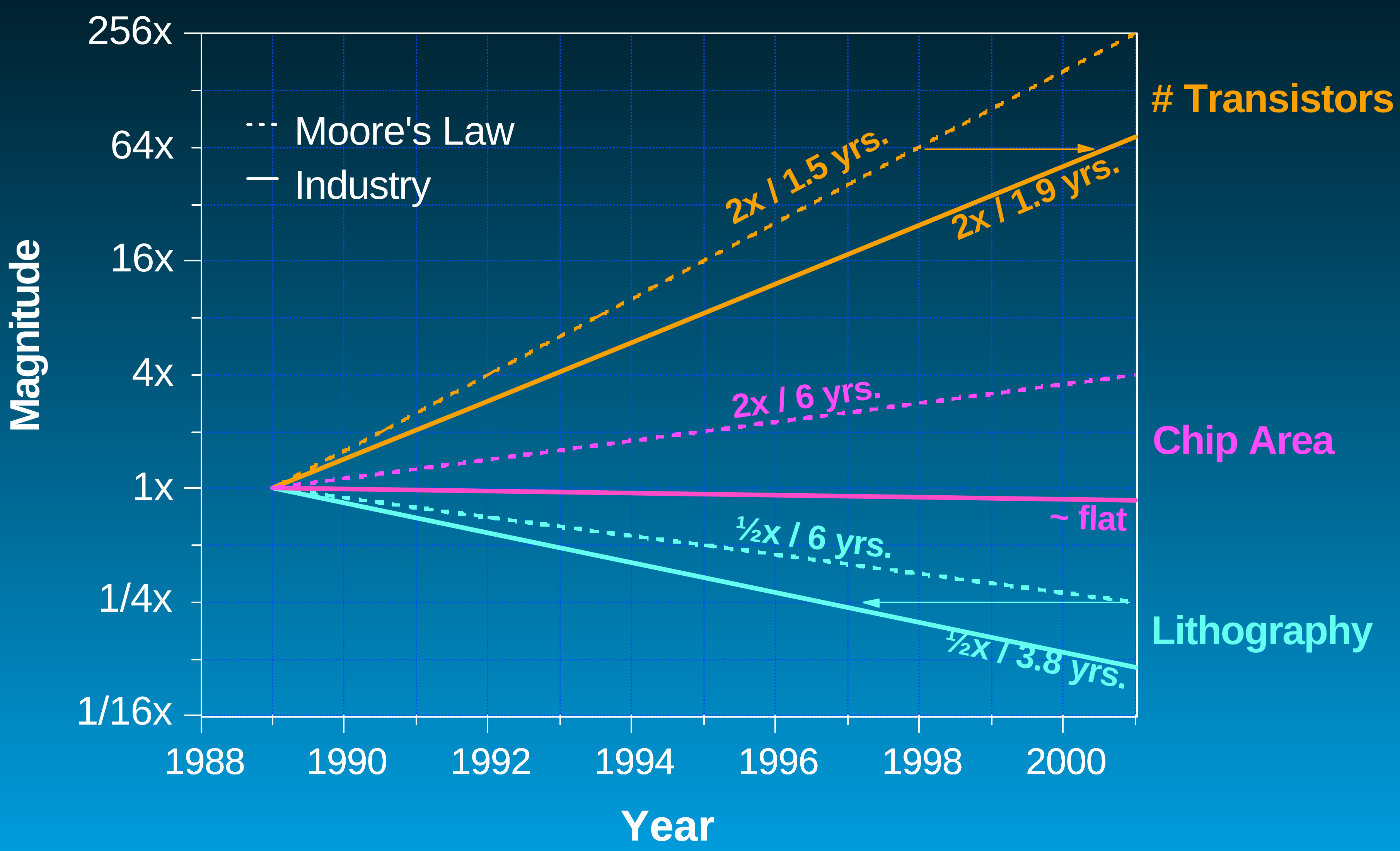
System Level Performance Improvement

Overall System Level Performance Improvement Will Come From Many Small Improvements



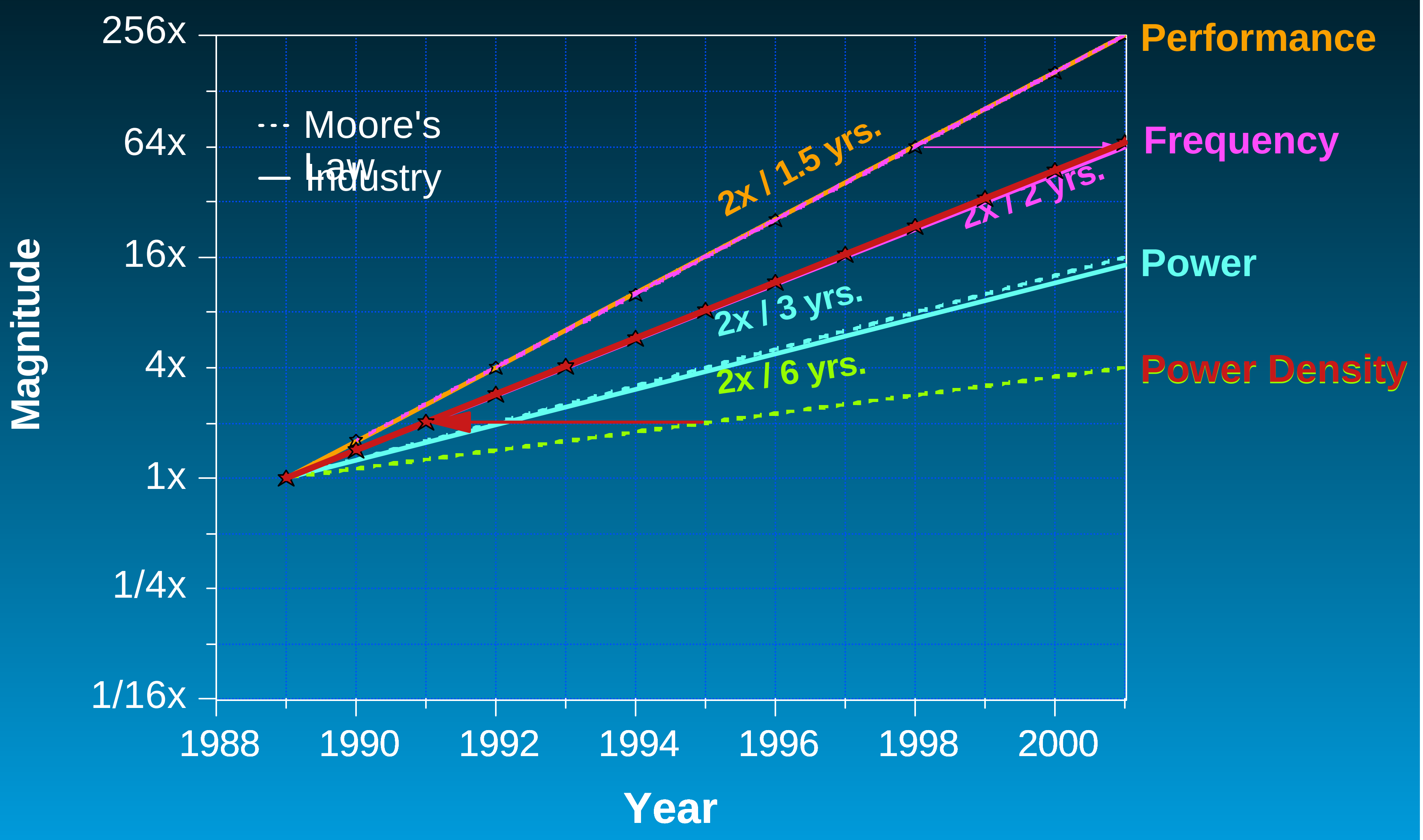

Moore's Law continues beyond conventional scaling
Power becomes the limiting metric
The integration focus moves from circuit to processor

Microprocessor Size Trends



IBM

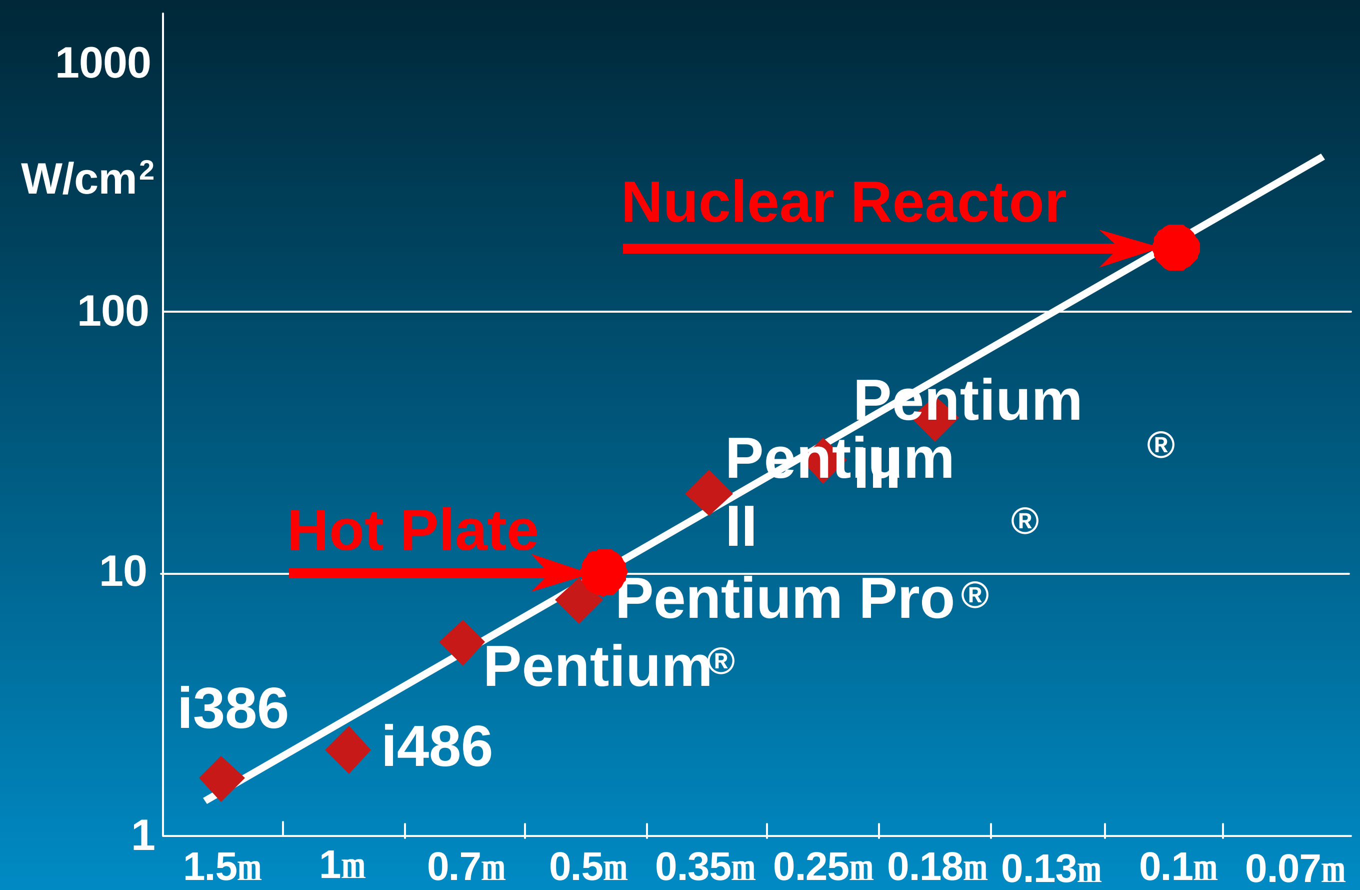
Microprocessor Performance Trends



Microprocessor Scaling Trends

Processor	486DX	Device Scaling	Moore's Law	Pentium 4
Date	04/10/89	2001	2001	04/23/01
Technology (um)	1	0.25	0.25	0.18
Vdd (V)	5	1.25	1.25	1.75
FPG	5	20	20	51
Frequency (MHz)	25	100	6400	1700
SpecInt95	0.5	2.0	128	71
# Transistors (M)	1.2	1.2	307	42
Chip Size (sq. mm)	165	10	660	216
Power (W)	4	0.25	66	64
Power Density (W/sq. cm)	2.5	2.5	10	29.5

Power Density: The Fundamental Problem



Source: Fred Pollack, Intel. New Microprocessor Challenges in the Coming Generations of CMOS Technologies, Micro32

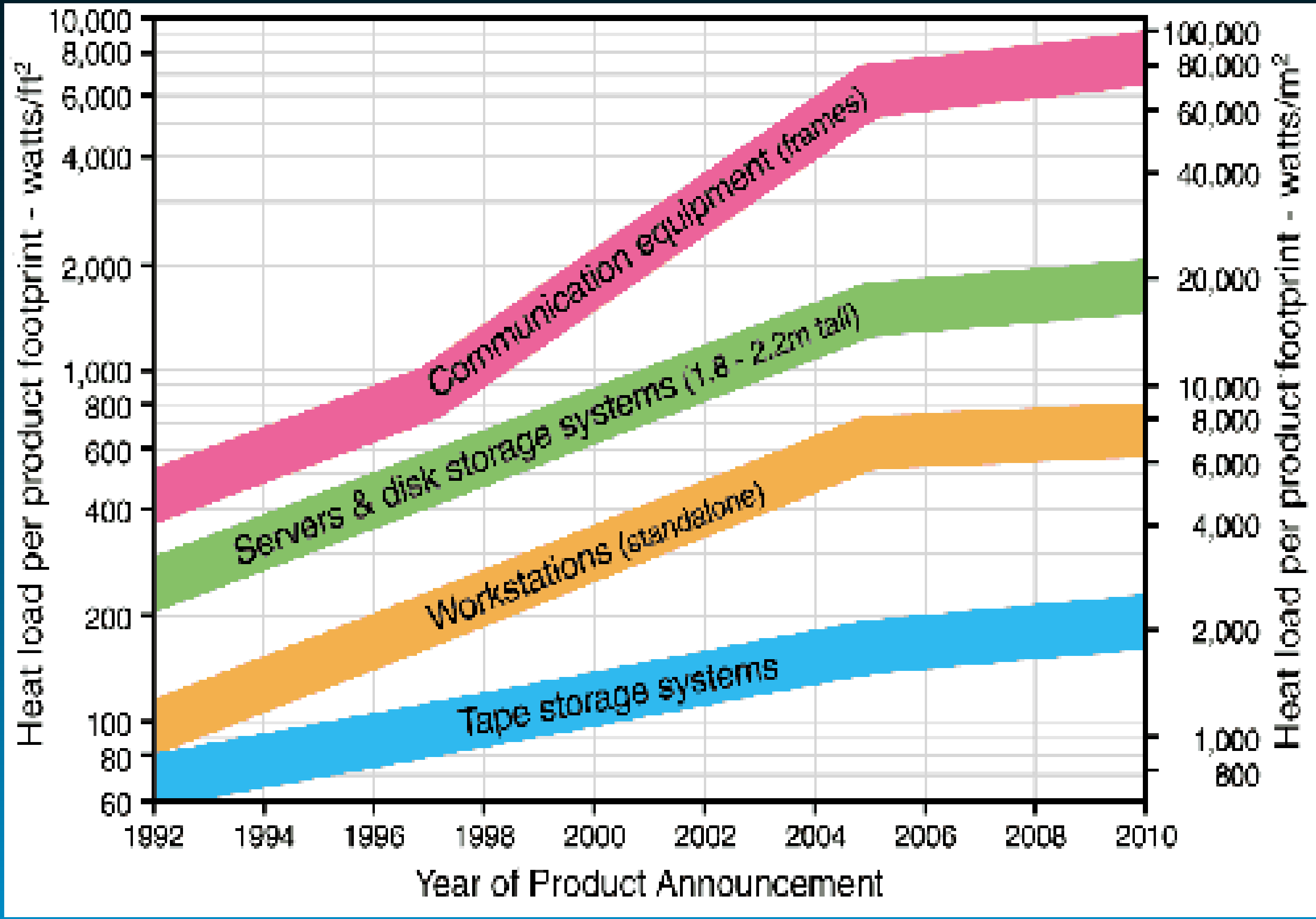
IBM

Power

IT electrical power needs are projected to reach crisis proportions

- Server farm energy consumption is increasing exponentially
  - ...more Watts/sq. ft than semiconductor or automobile plants
  - ...energy needs constitute 60%of cost
- Interesting anecdotes
  - The "2,400 megawatt problem":
    - 27 farms proposed for *South King County* will require as much energy as *Seattle* (including *Boeing*)
  - *Exodus* considering building power plant near its *Santa Clara* facility
  - *San Jose City Council* approved 250 MW power plant for *US DataPort* server farm
    - and installation of 80 back-up diesel generators

Server Farm Heat Density Trend

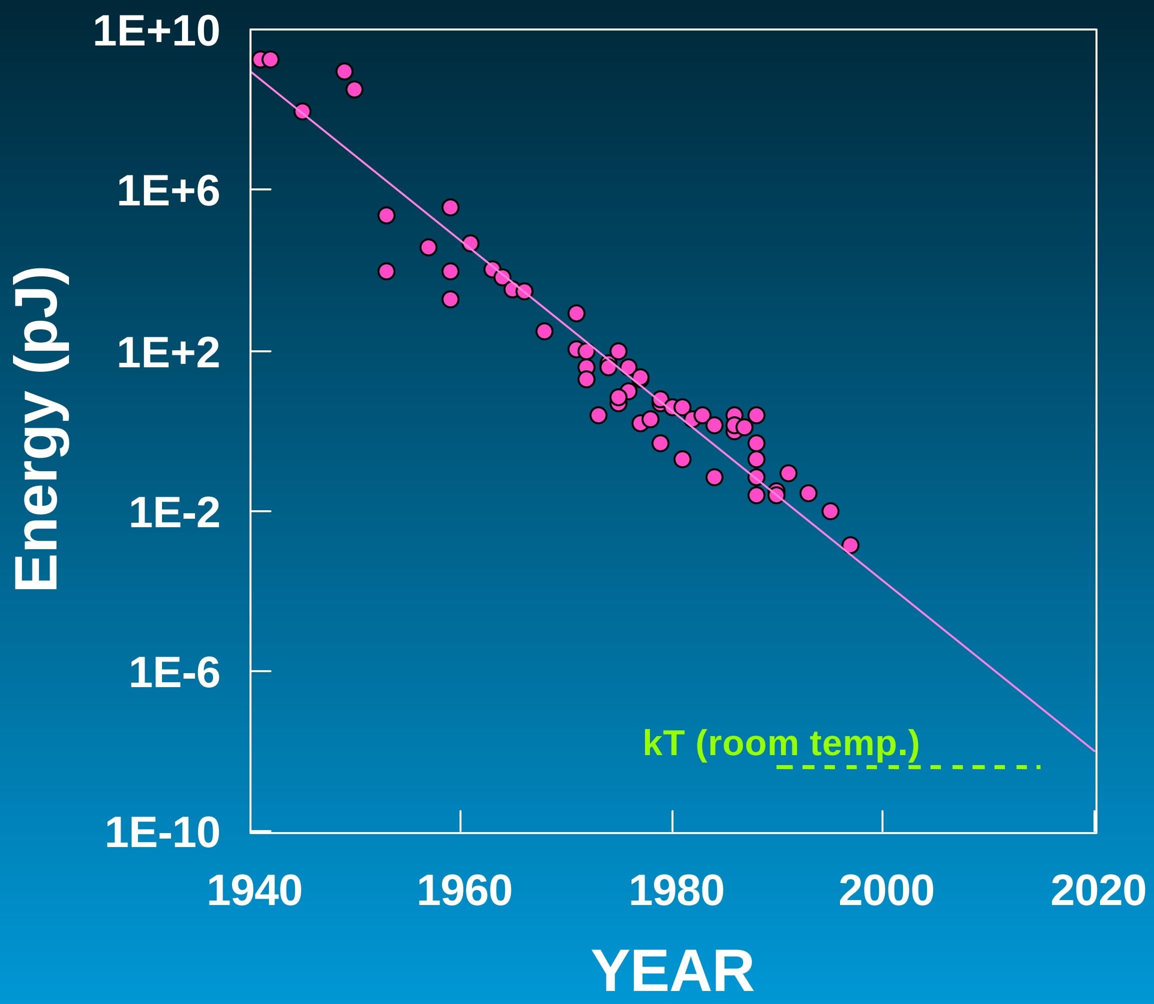


Highest Communication: 28% AGR

Lowest Tape storage: 7%

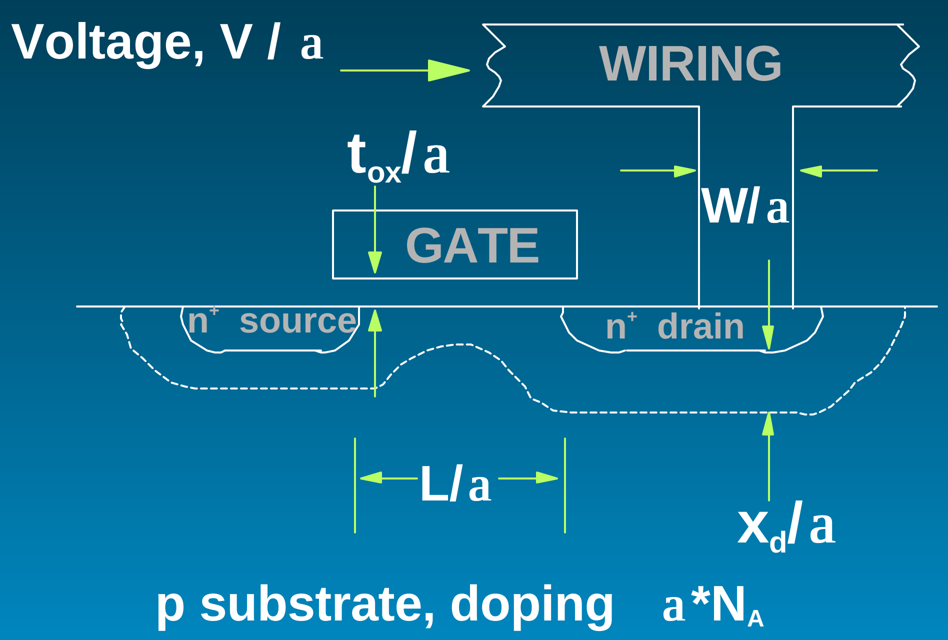
*** Slower growth after 2005 due to improvement in semiconductor power consumption**

Energy Dissipated per Logic Operation



Device Scaling

Scaled Device



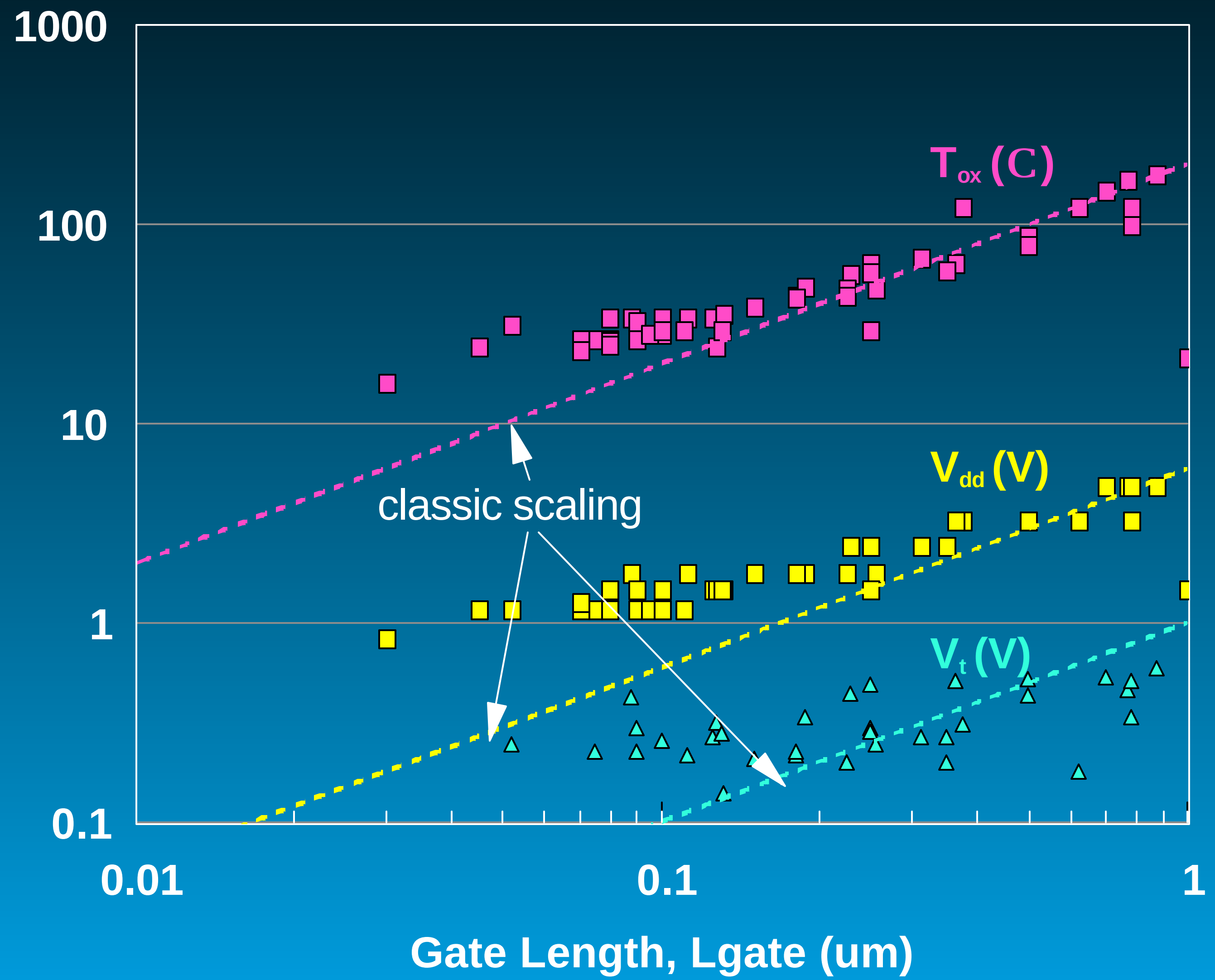
SCALING:

Voltage:	V/ a
Oxide:	t_{ox} / a
Wire width:	W/ a
Gate width:	L/ a
Diffusion:	x_d / a
Substrate:	a * N_A

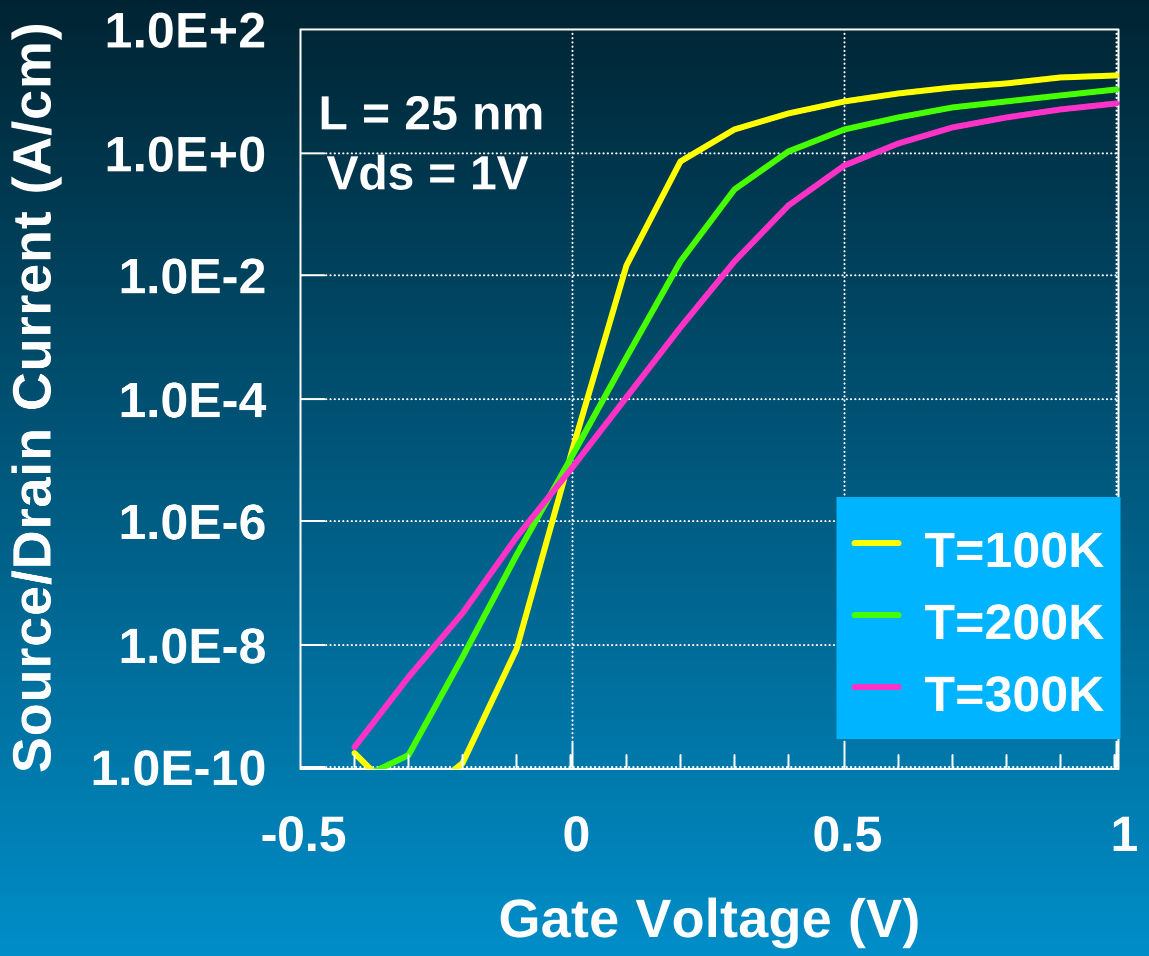
RESULTS:

Higher Density:	~ a^2
Higher Speed:	~ a
Lower Power/ckt:	~1/ a^2
Power Density:	~Constant

MOSFET Device Parameter Trends

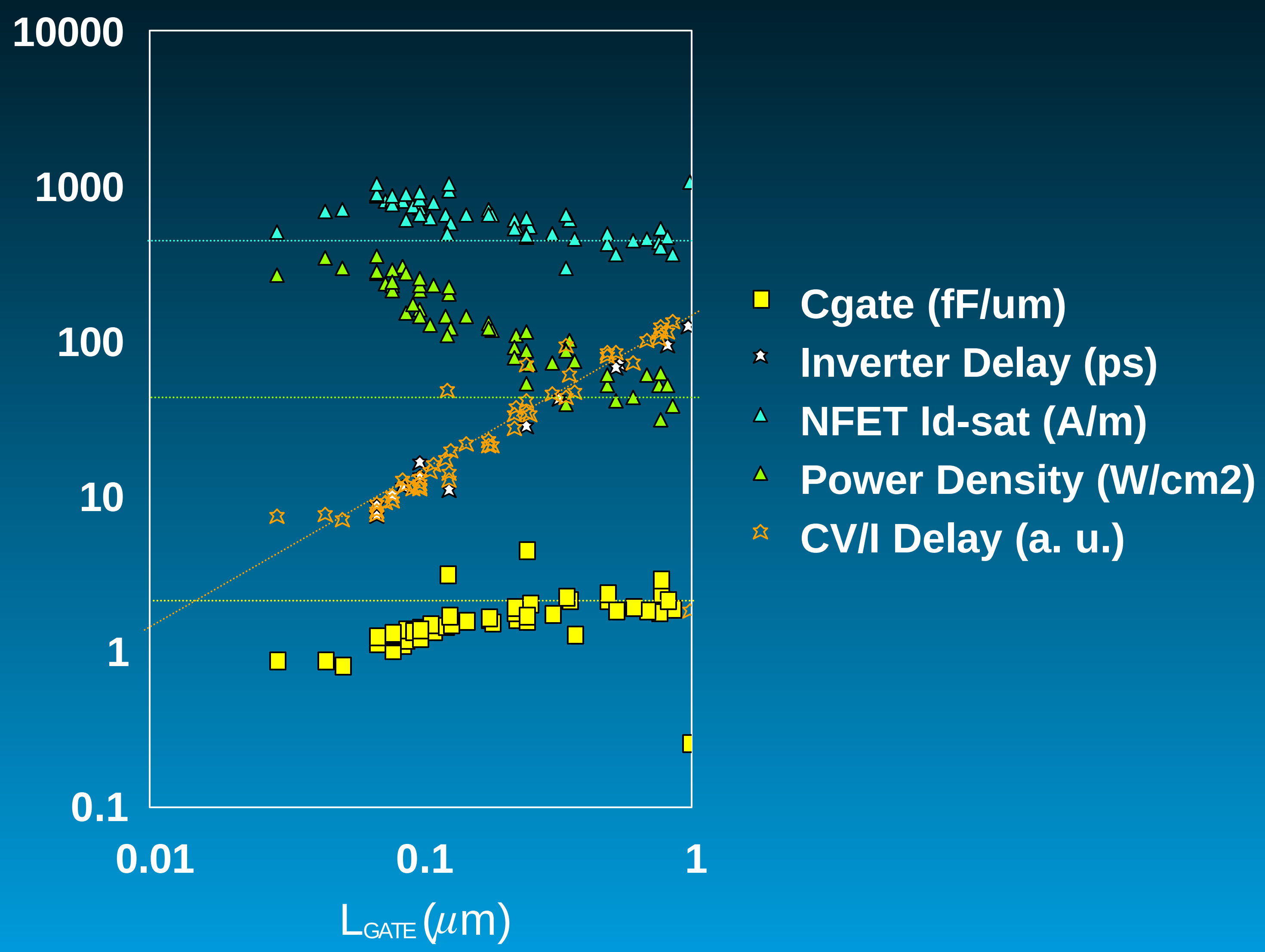


Low Temperature CMOS

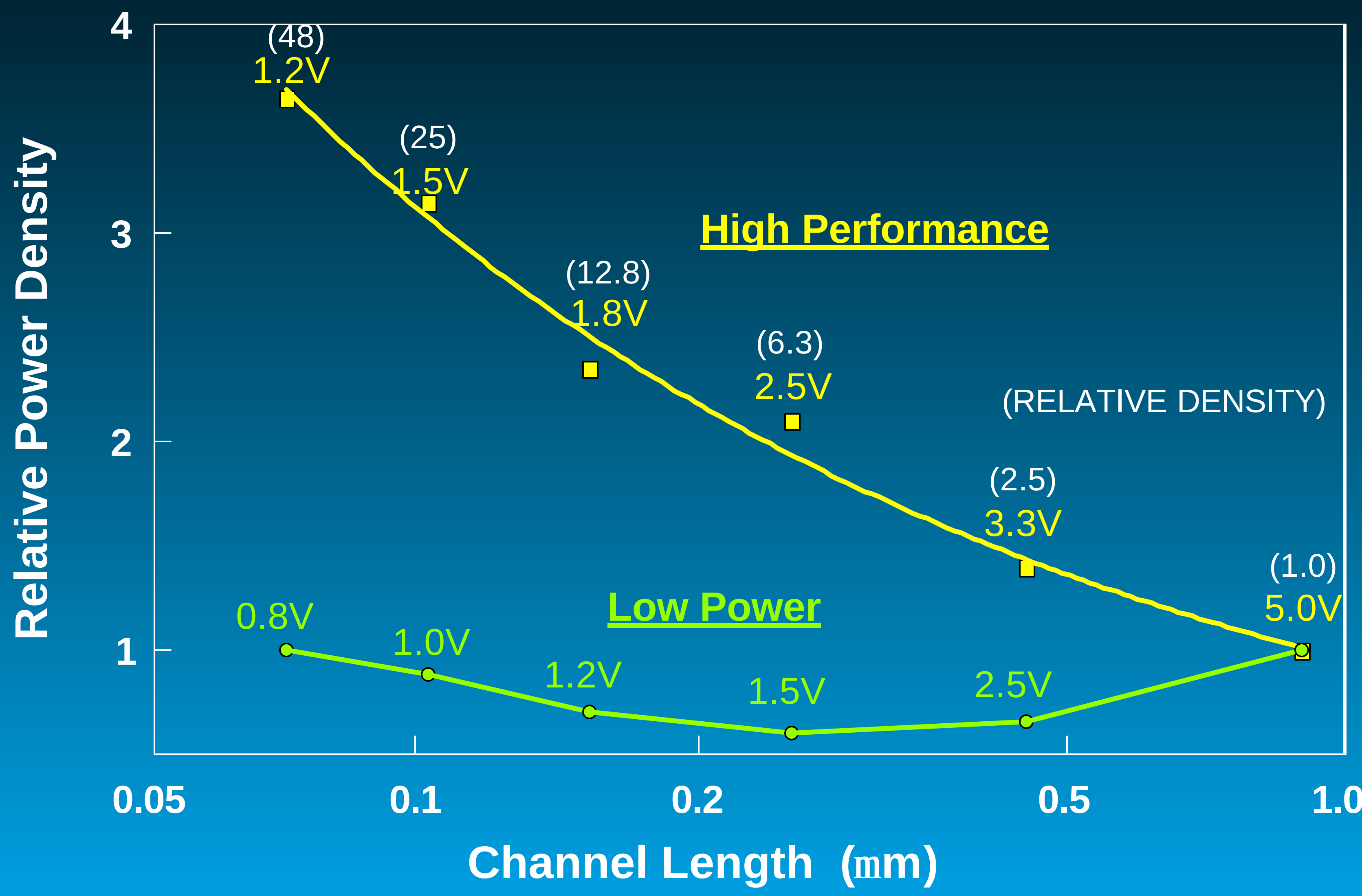


- Subthreshold slope steepens as temperature is reduced

CMOS Performance Parameter Trends



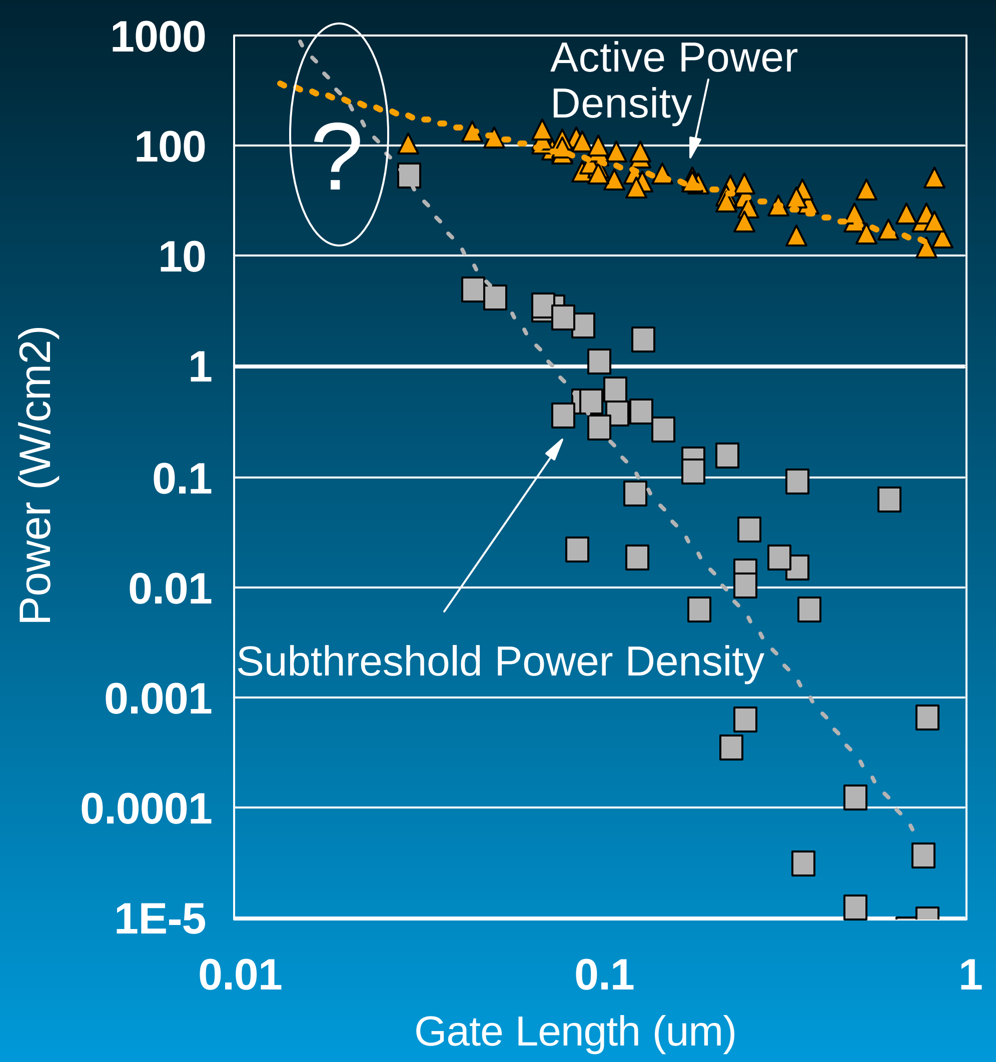
Relative Power Density in Scaled CMOS



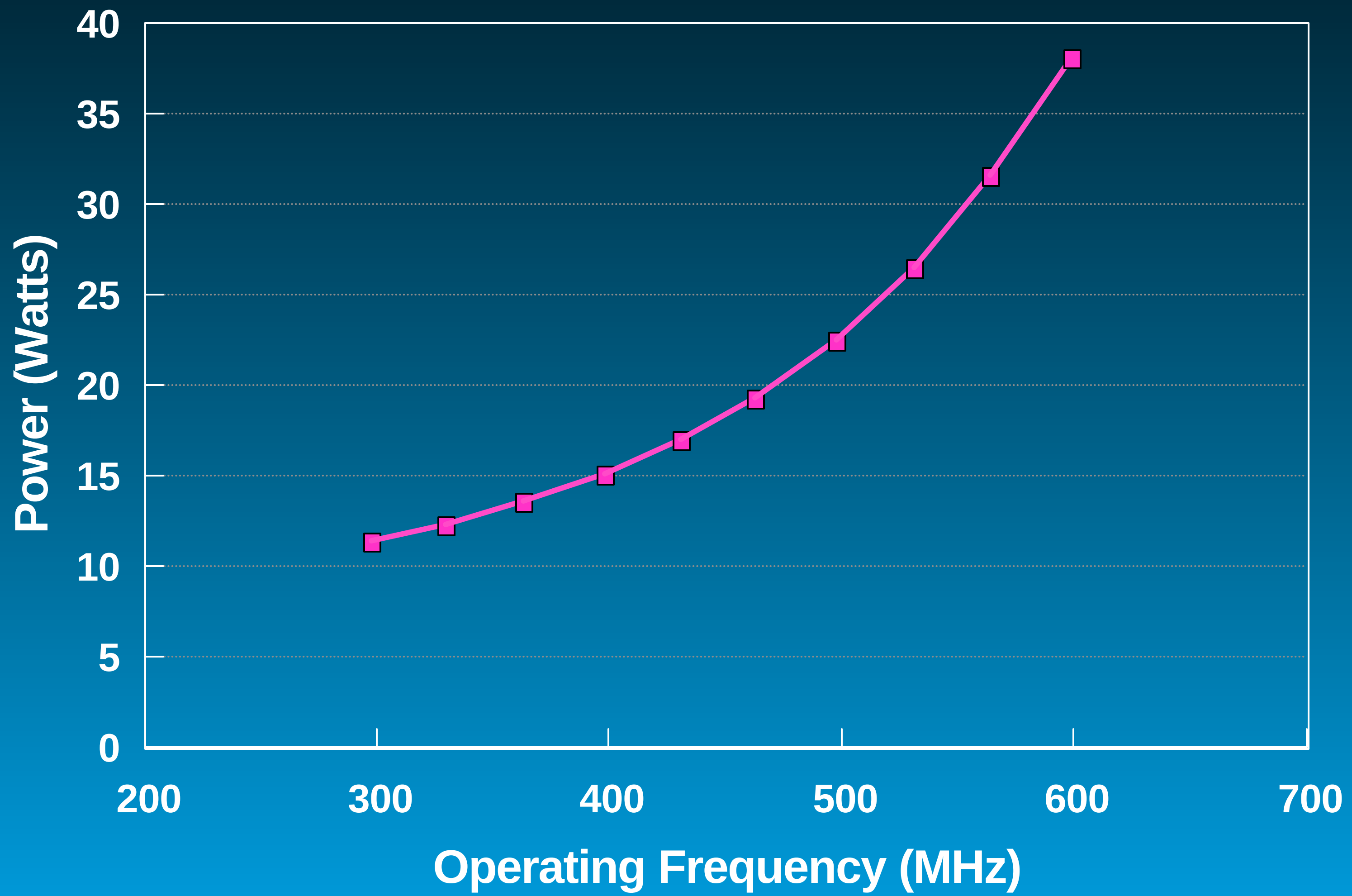
After B. Davari, et al., IEEE Proc. Vol. 83, p. 595, 1995.

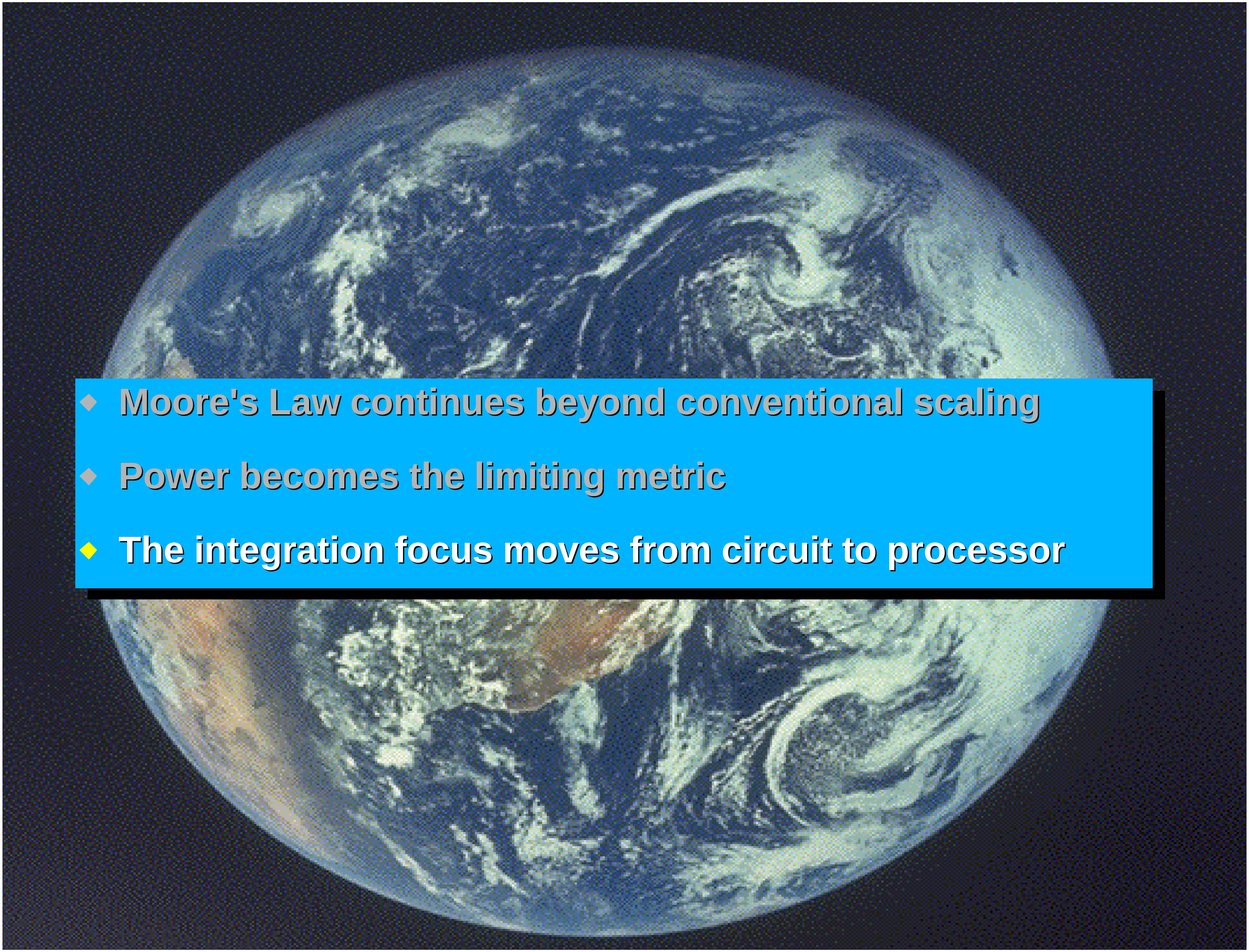
IBM

CMOS Power Density Trends

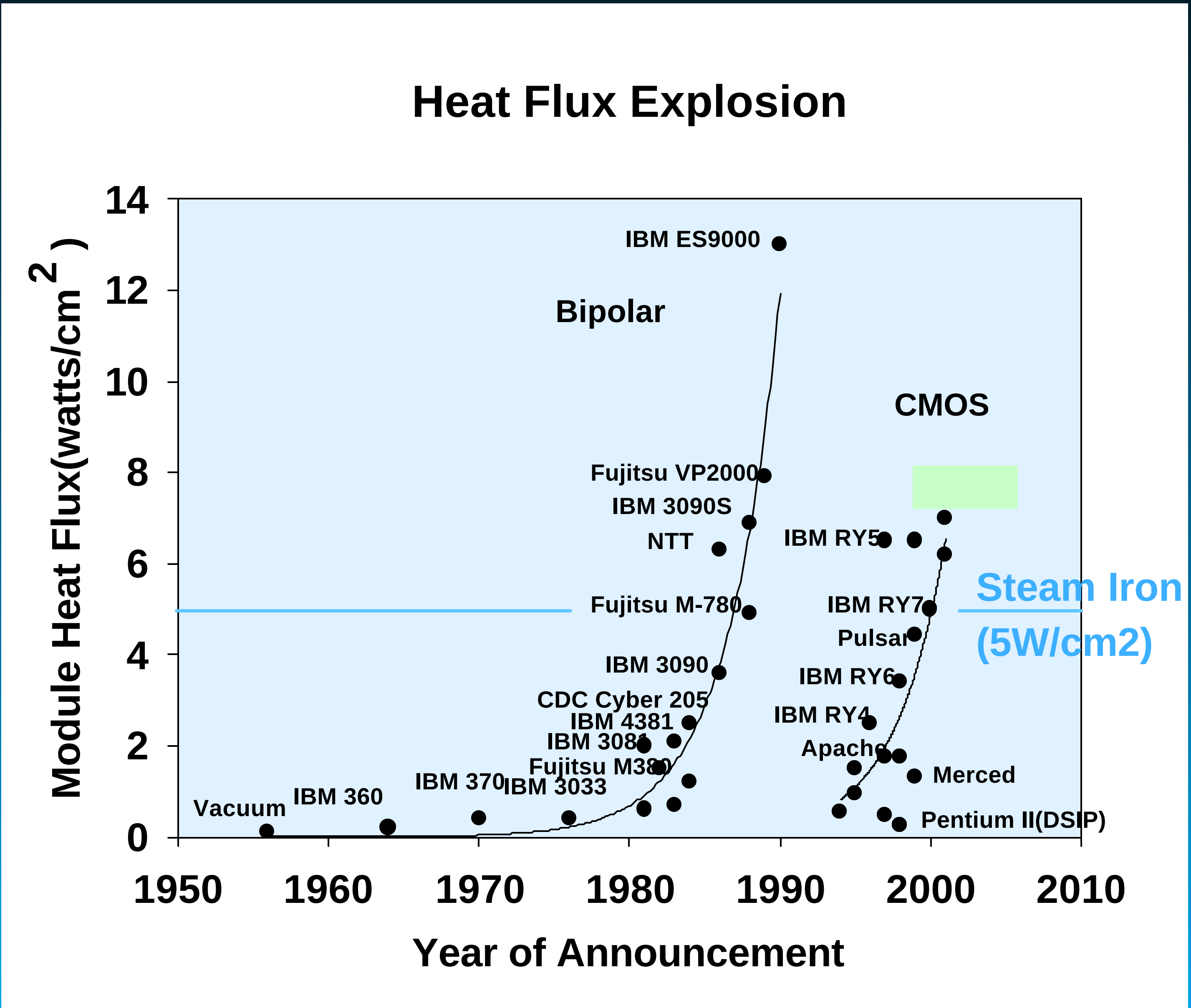


Microprocessor Power Draw vs. Frequency

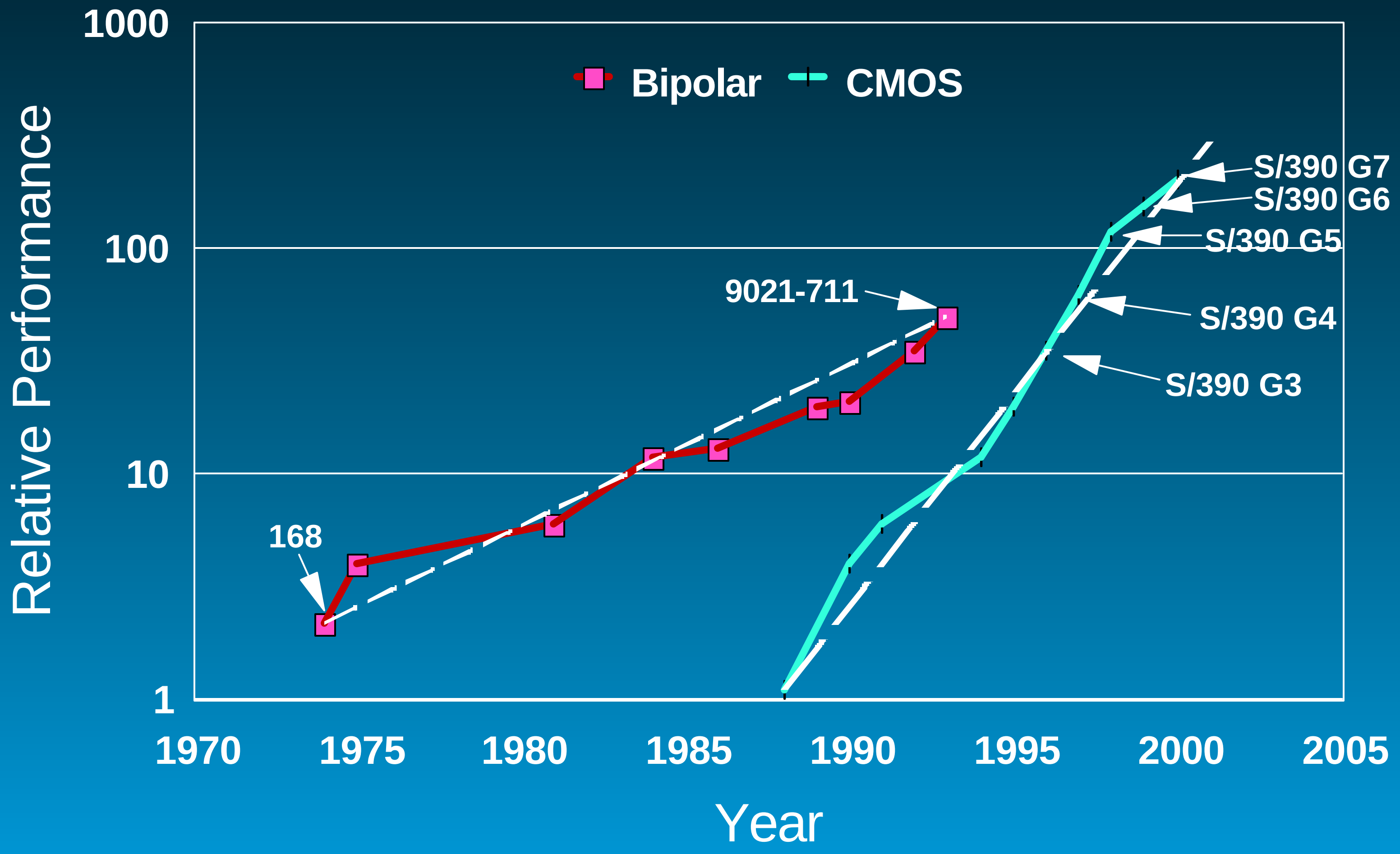



Moore's Law continues beyond conventional scaling
Power becomes the limiting metric
The integration focus moves from circuit to processor

We've been here before!



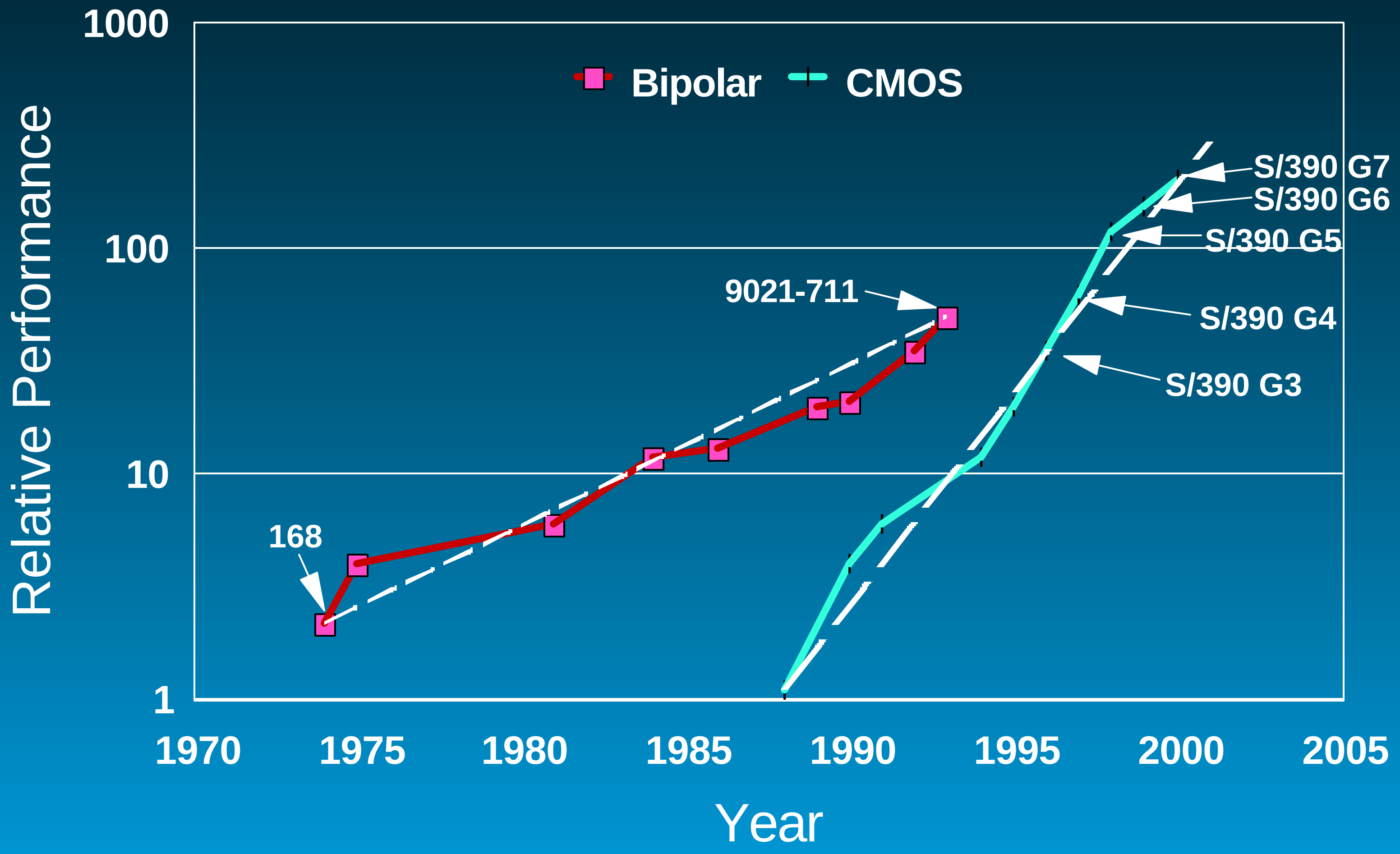
S/390 Mainframe CPU Performance



S/390: Comparison of Bipolar and CMOS

	ES9000 9X2	S/390 G5
Technology	Bipolar	CMOS
Total Chips	5000	29 (12 CPUs)
Total Parts	6659	92
Weight (lbs)	31.1 K	2.0 K
Power Req (KW)	153	5
Chips/processor	390	1
Maximum Memory (GB)	10	24
Space (sq ft)	672	52

S/390 Mainframe CPU Performance

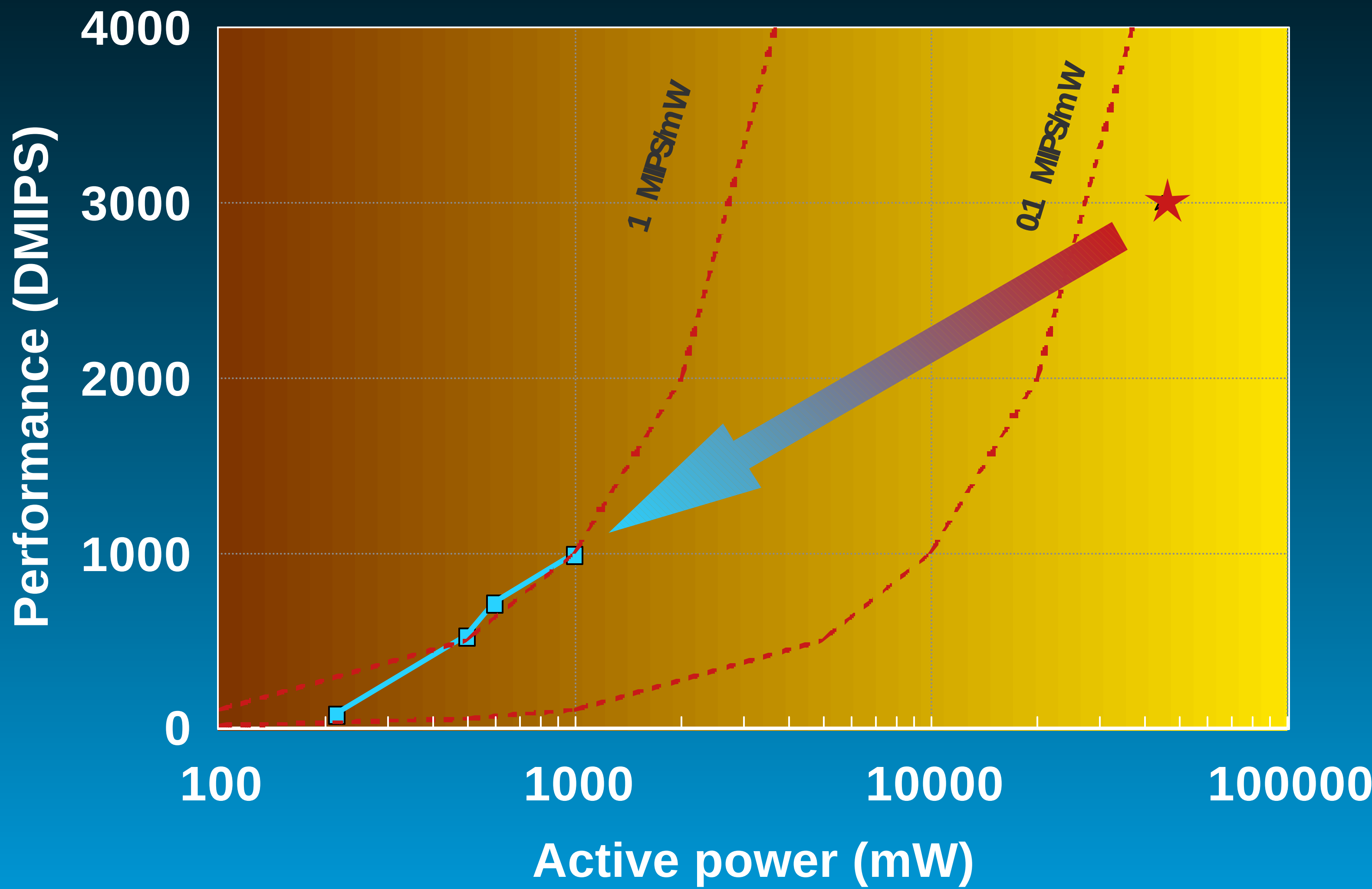


Focus on massively parallel systems

- **Use slower processors with much greater power efficiency**
- **Scale to desired performance with parallel systems**
- **Workload scaling efficiency must sustain power efficiency**
- **Physical distance must be small to keep communication power manageable.**

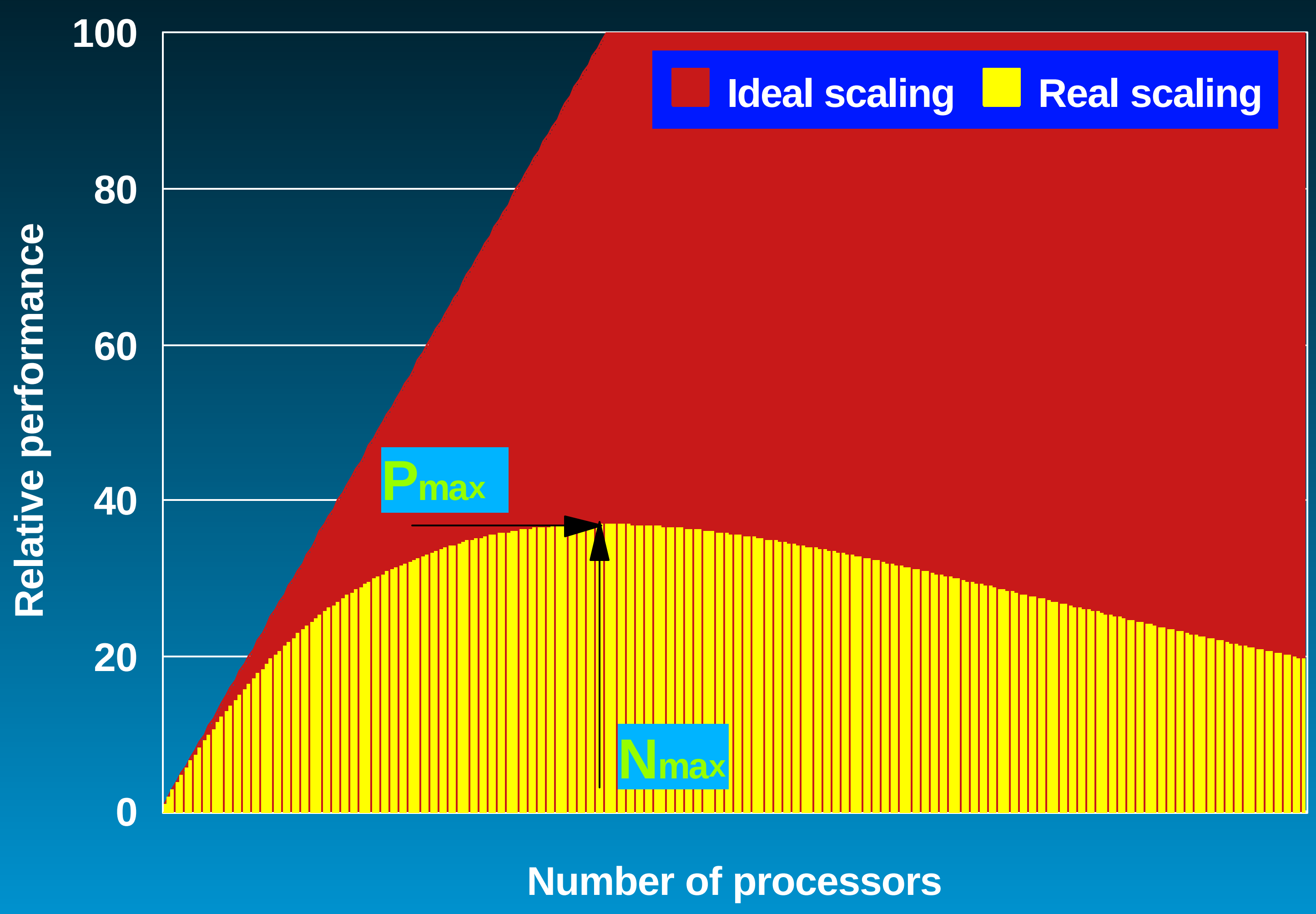
Example: Processor A is slower than B by a factor S but more power efficient by E. Then MP System A at the same performance as MP System B has lower power by E/S.

Microprocessor Efficiencies

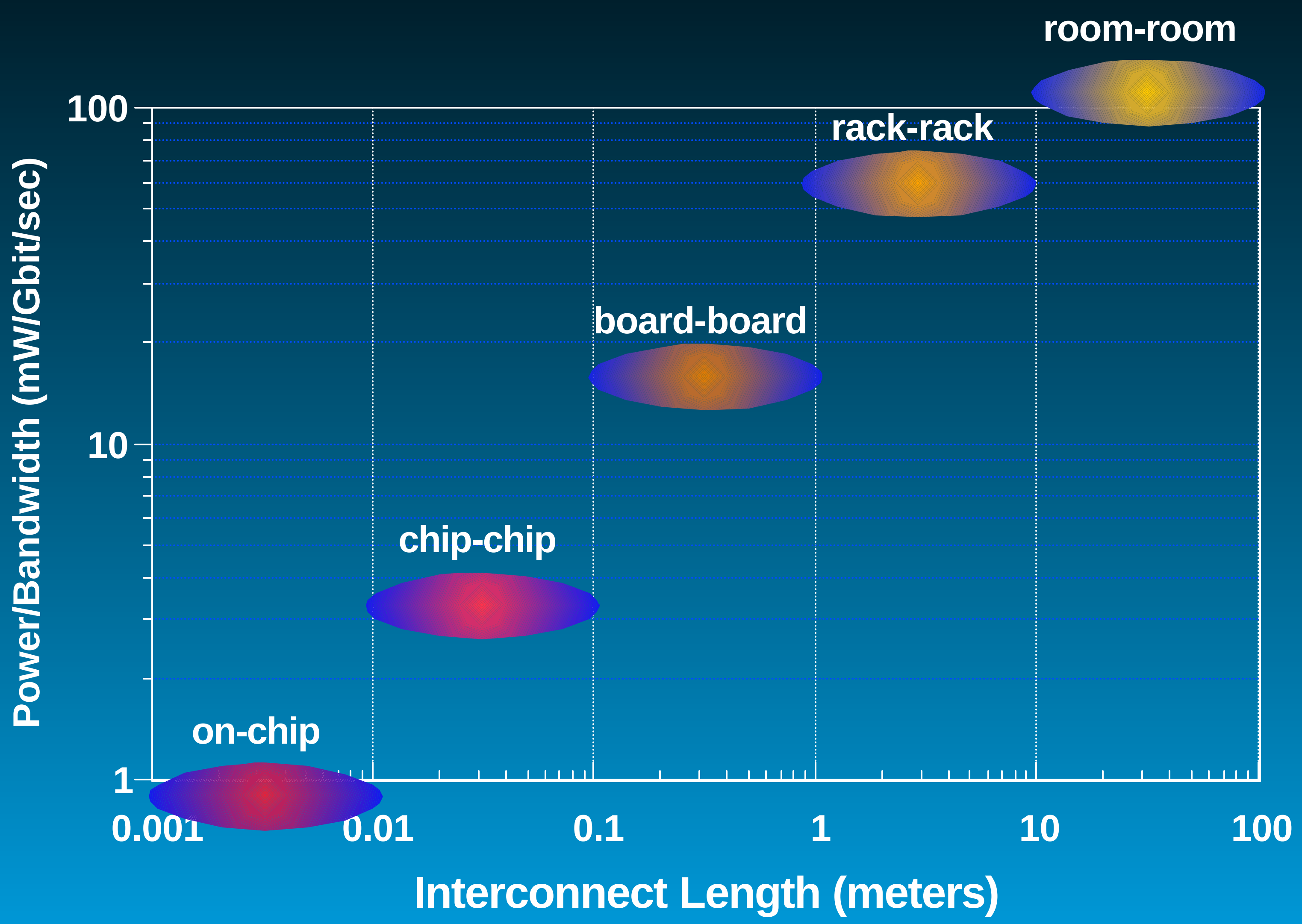


IBM

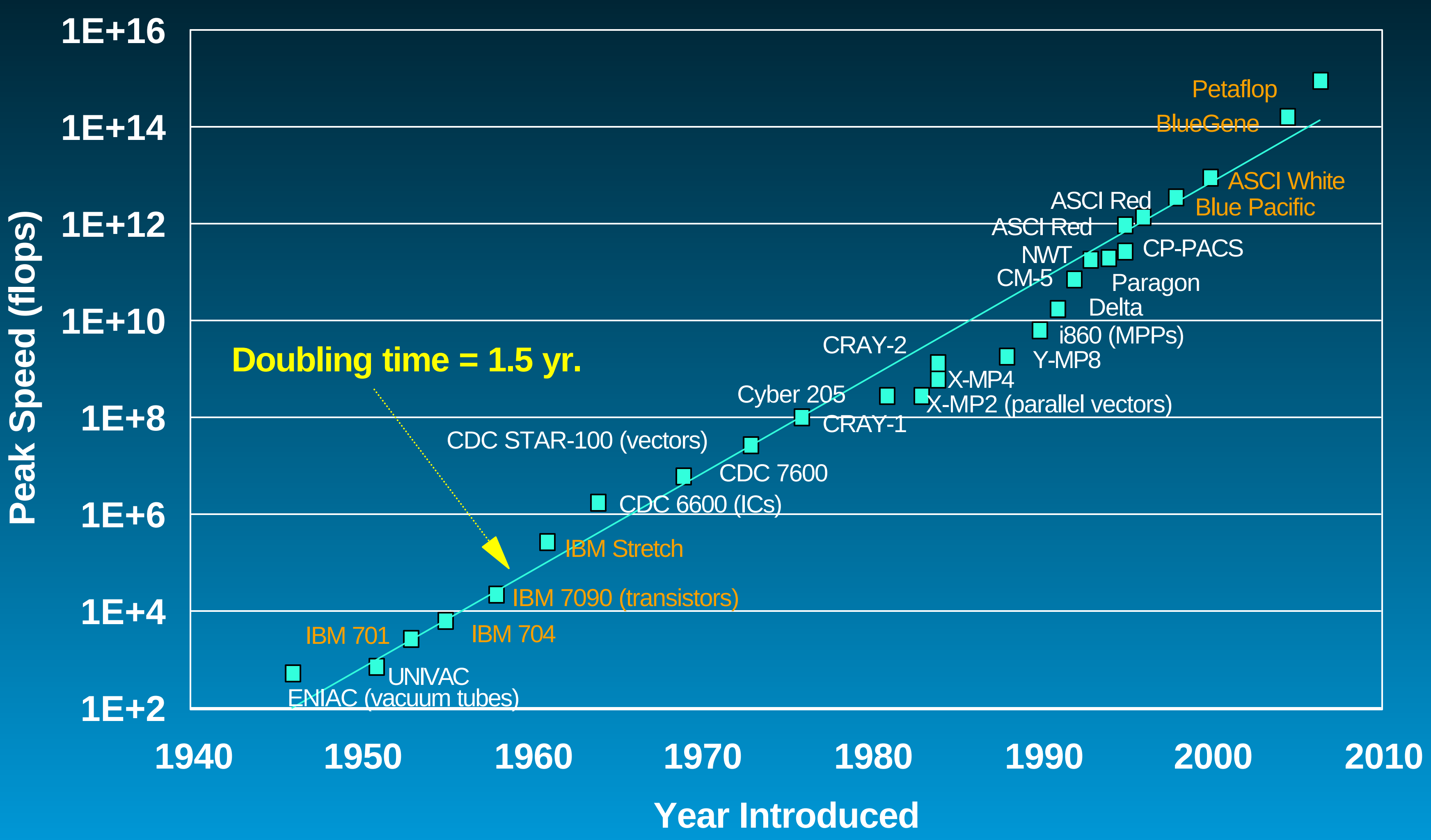
Parallel Performance Scaling Model



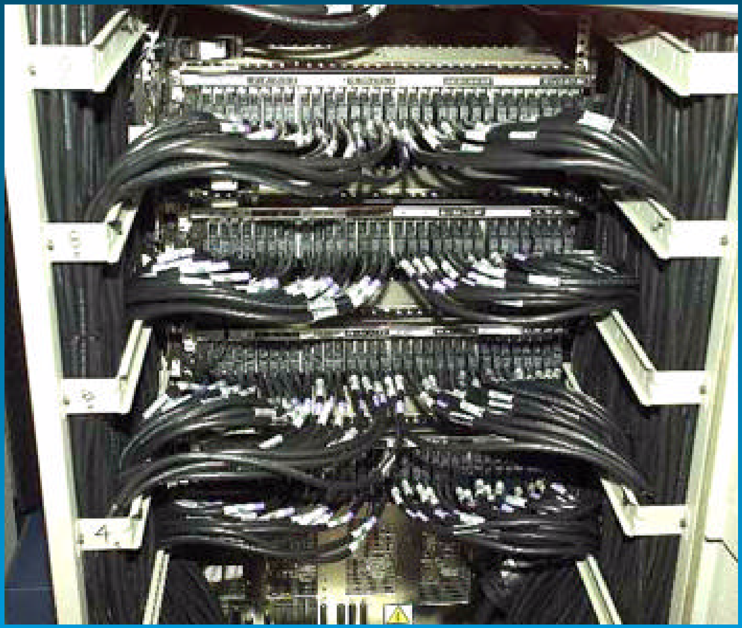
Power / Bandwidth by Interconnect Length

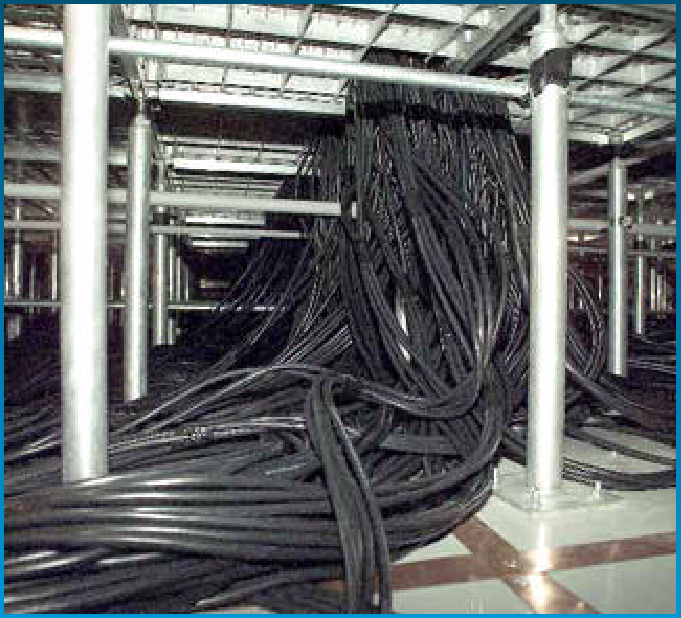


Supercomputer Peak Performance

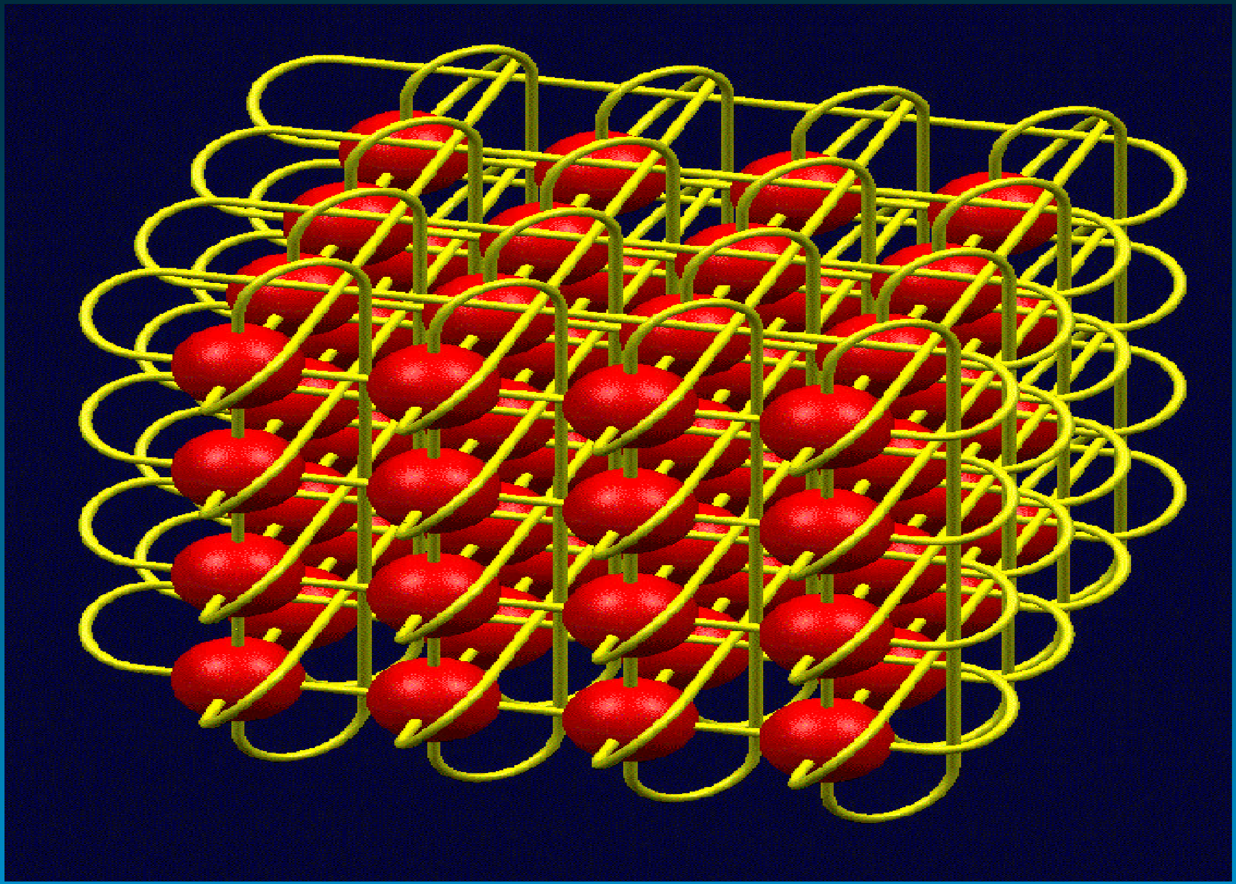


ASCI White



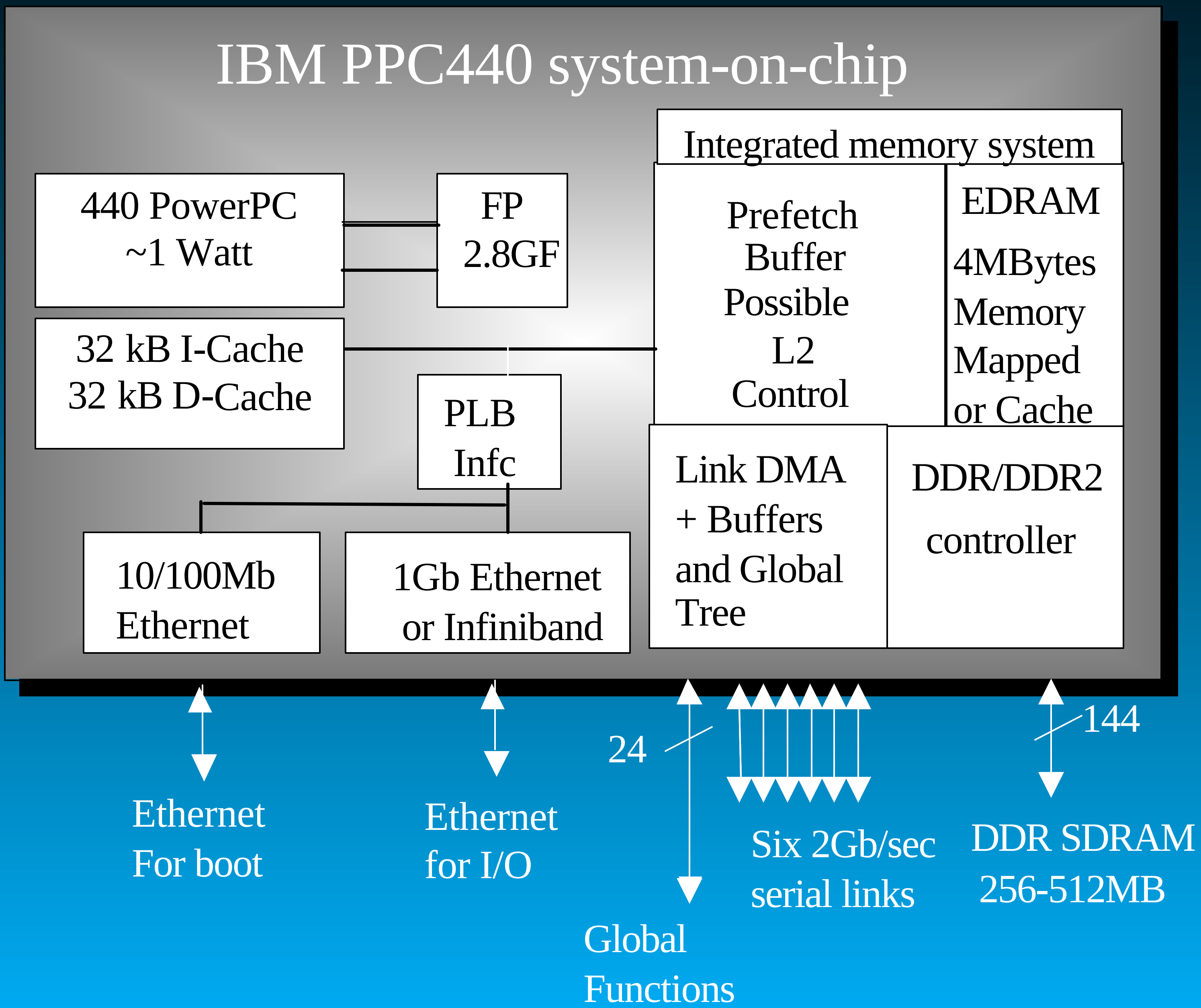


Cellular Architecture



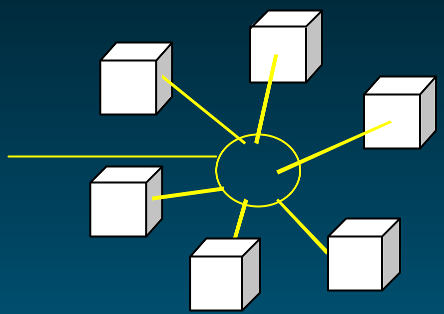
- computational efficiency ~ 0.2 GFLOP/W

Example of a Cellular Node



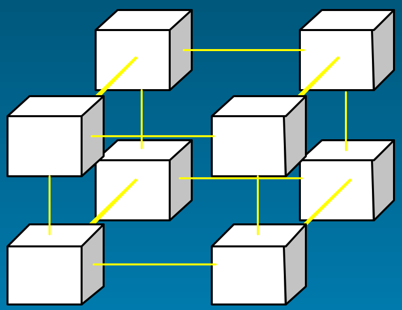
Cellular Communication Networks

- 65536 nodes interconnected with three integrated networks



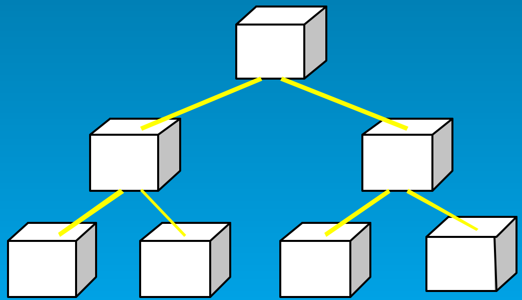
Ethernet

- Incorporated into every node ASIC
- Disk I/O
- Host control, booting and diagnostics



3 Dimensional Torus

- Virtual cut-through hardware routing to maximize efficiency
- 2.8 Gb/s on each of 12 node links (total 4.2 GB/s per node)
- Communication backbone
- 134 TB/s total torus interconnect bandwidth
- 1.4/2.8 TB/s bisectional bandwidth

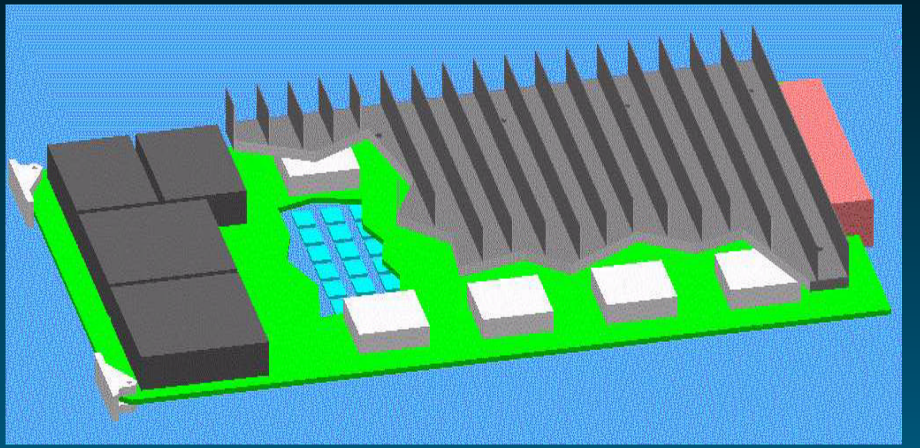


Global Tree

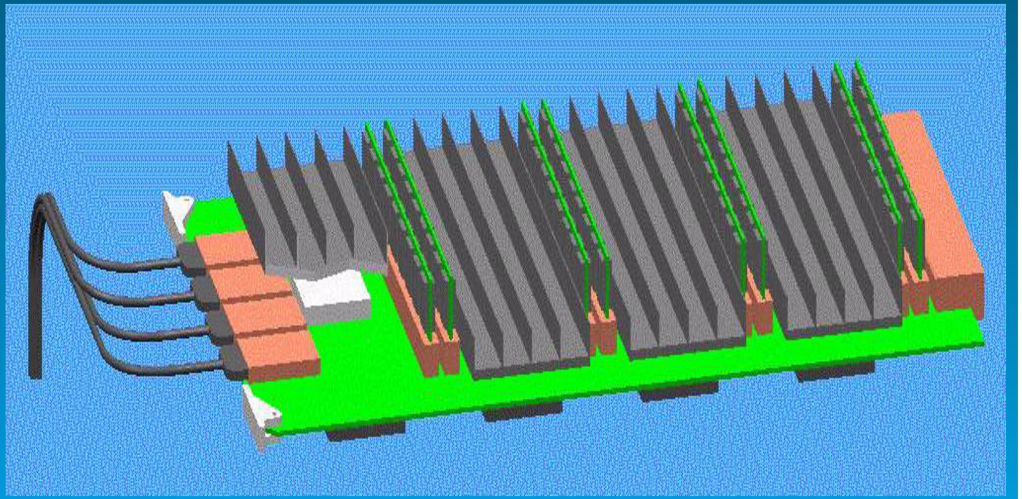
- One-to-all or all-all broadcast functionality
- Arithmetic operations implemented in tree
- ~1.4 GB/s of bandwidth from any node to all other nodes
- Latency of tree less than 1usec
- ~90TB/s total binary tree bandwidth (64k machine)

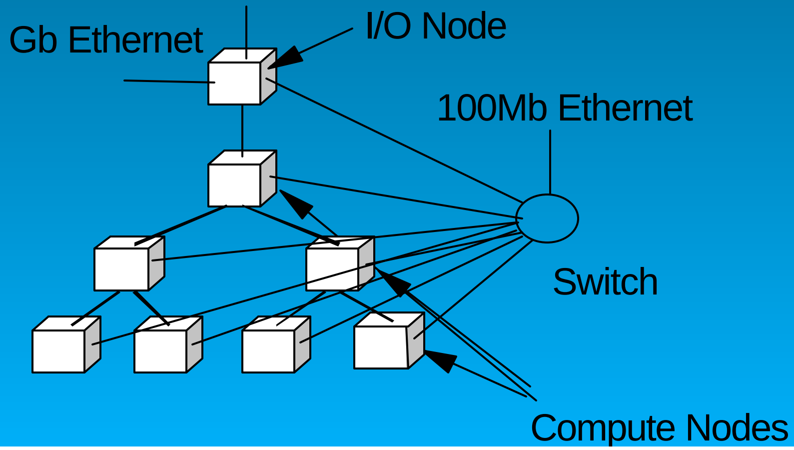
Node Card and I/O Card Design

- **Compute cards**
  - 8 processors, 2 x 2 x 2 (x,y,z)
  - 256 MB RAM each processor
  - Redundant power supplies
  - Fast Ethernet



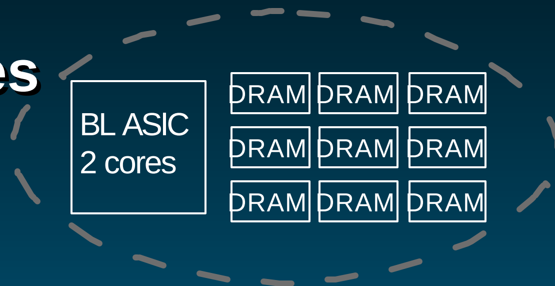
- **I/O cards**
  - 4 processors (no torus)
  - 512MB-1GB each processor
  - Redundant Power Supplies
  - Fast and 1Gb Ethernet





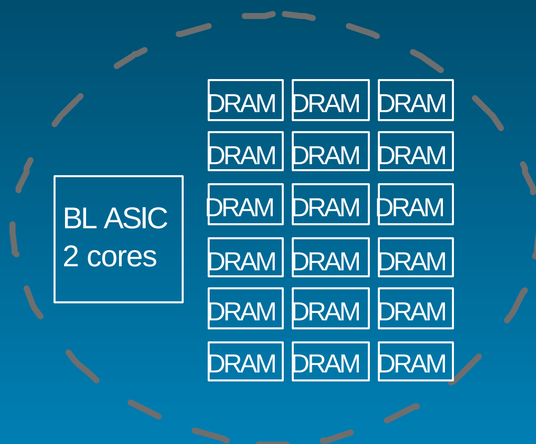
Rack Design

- **1024 compute nodes**
  - **256 GB DRAM**
  - **2.8TF peak**



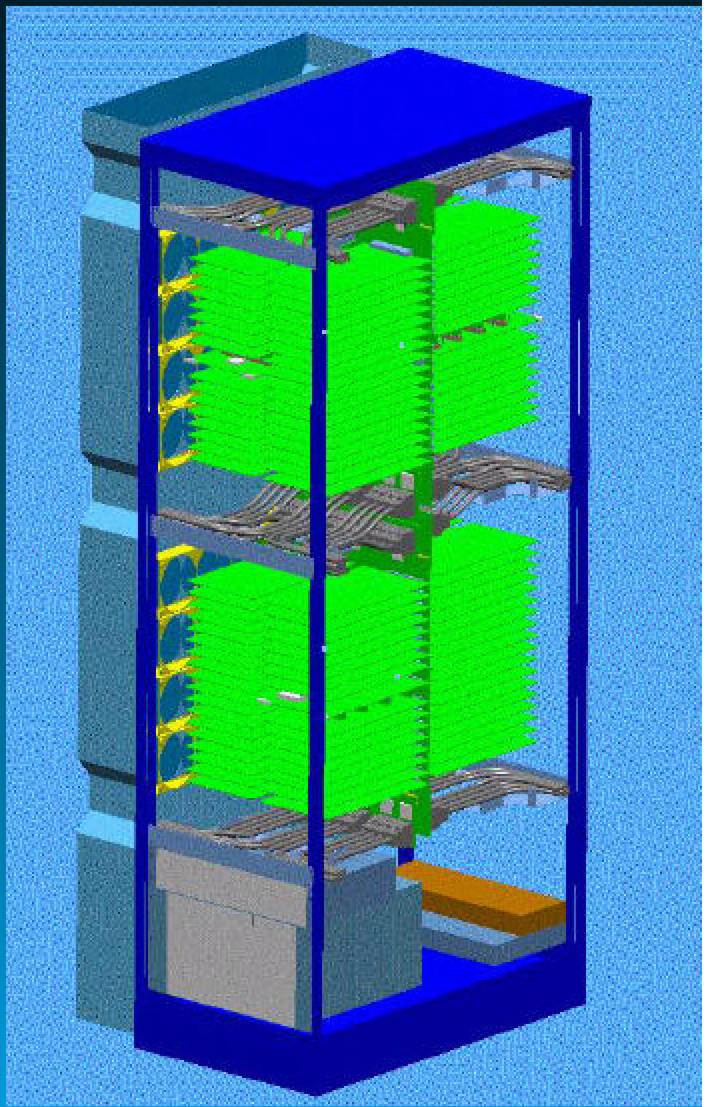
One compute node

- **16 I/O nodes**
  - **8 GB DRAM**
  - **16 Gb Ethernet**



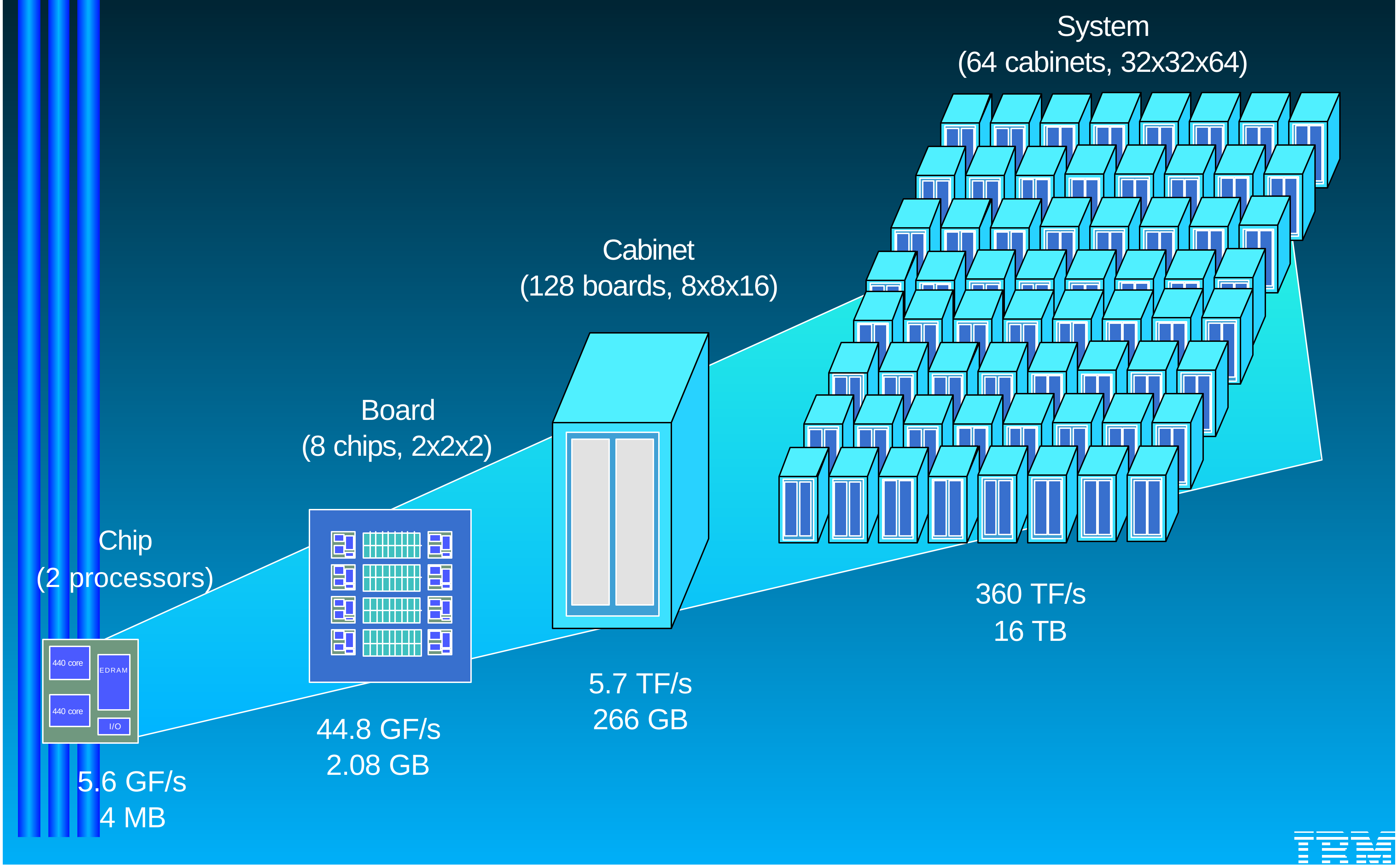
One I/O node

- **~15 KW, air cooled**
  - ✓ **1+1 or 2+1 redundant power**
  - ✓ **2+1 redundant fans**



IBM

Building a Cellular System



- Moore's Law continues beyond conventional scaling
  - Technology innovation will overcome limits
- Power becomes the limiting metric
  - Technology trend is to higher power density
- The integration focus moves from circuit to processor
  - Radical power reduction depends on efficient processors
  - Massively parallel systems have great potential

(Hopefully Not) The End!