

**IFIP Working Group 10.3 on Concurrent Systems**



# *An Overview of High Performance Computing*

---

**Jack Dongarra  
University of Tennessee  
and  
Oak Ridge National Laboratory**



# Overview

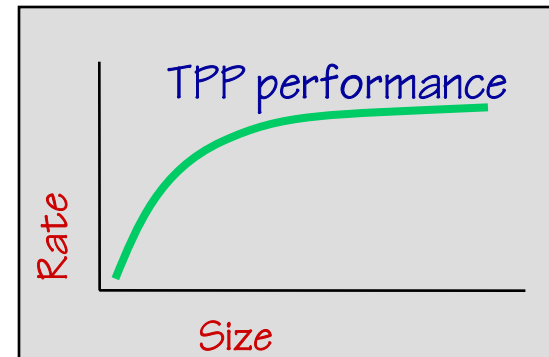
---

- ◆ **Look at fastest computers**
  - **From the Top500**
- ◆ **Some of the changes that face us**
  - **Hardware**
  - **Software**
  - **Algorithms**

## H. Meuer, H. Simon, E. Strohmaier, & JD

- Listing of the 500 most powerful Computers in the World
- Yardstick: Rmax from LINPACK MPP

$$Ax=b, \text{ dense problem}$$



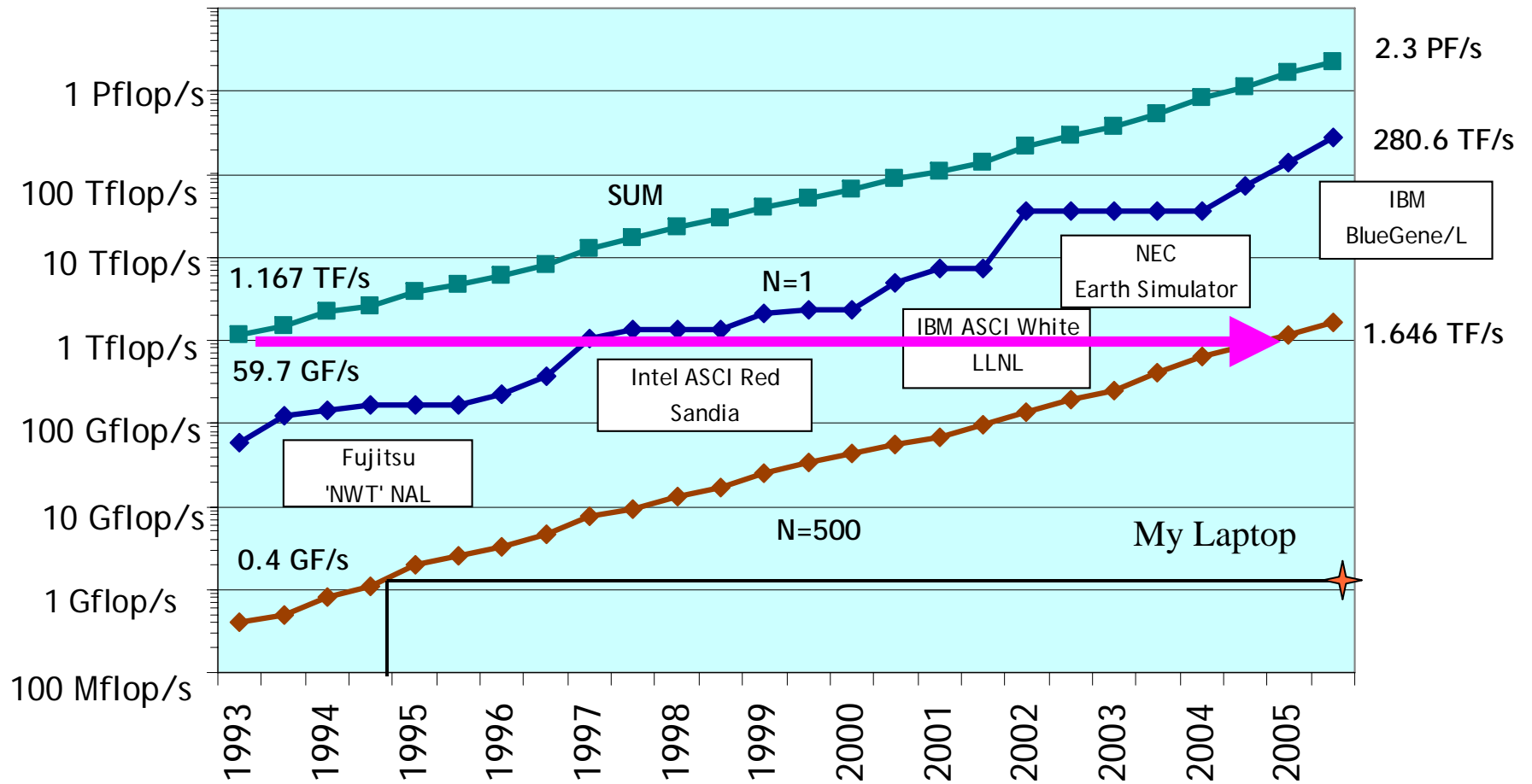
- Updated twice a year
- SC'xy in the States in November
- Meeting in Germany in June
- Started in 1993, 13 years of data

00

- All data available from [www.top500.org](http://www.top500.org)



# Performance Development



# Architecture/Systems Continuum

Tightly  
Coupled

◆ **Custom processor with custom interconnect**

- Cray X1
- NEC SX-8
- IBM Regatta
- IBM Blue Gene/L

- ◆ Best processor performance for codes that are not "cache friendly"
- ◆ Good communication performance
- ◆ Simpler programming model
- ◆ Most expensive

◆ **Commodity processor with custom interconnect**

- SGI Altix
  - Intel Itanium 2
- Cray XT3, XD1
  - AMD Opteron

- ◆ Good communication performance
- ◆ Good scalability

◆ **Commodity processor with commodity interconnect**

- Clusters
  - Pentium, Itanium, Opteron, Alpha
  - GigE, Infiniband, Myrinet, Quadrics

- ◆ Best price/performance (for codes that work well with caches and are latency tolerant)
- ◆ More complex programming model

Loosely  
Coupled

- NEC TX7
- IBM eServer
- Dawning

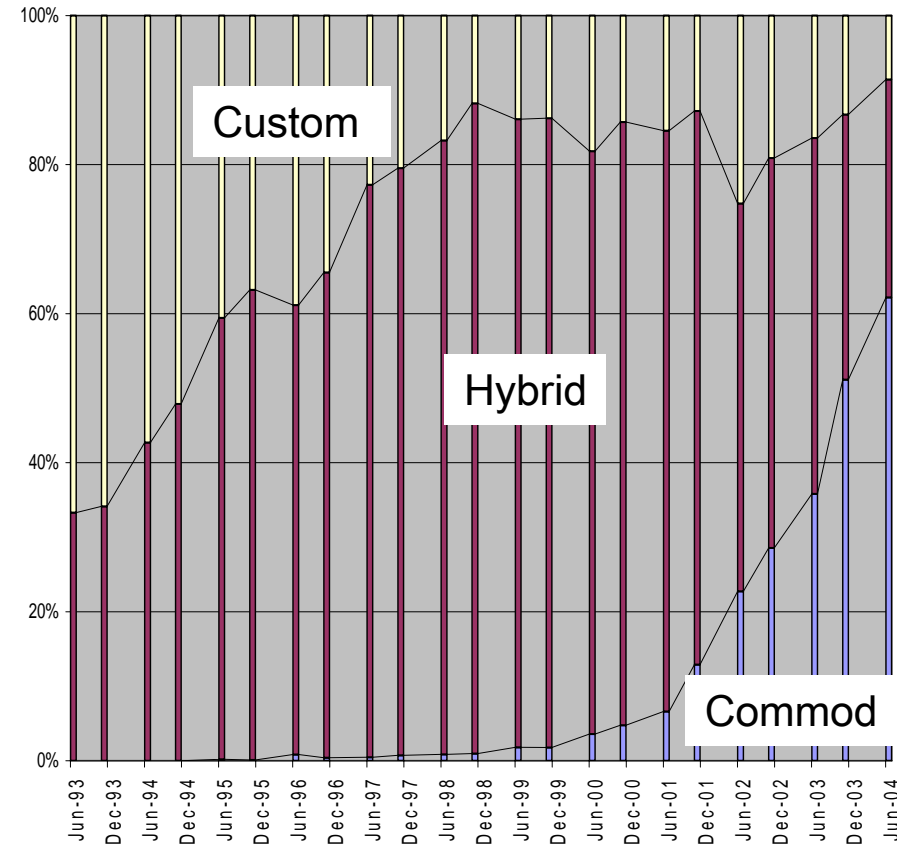
oo

# Architecture/Systems Continuum

Tightly  
Coupled

- ◆ **Custom processor with custom interconnect**
  - Cray X1
  - NEC SX-8
  - IBM Regatta
  - IBM Blue Gene/L
- ◆ **Commodity processor with custom interconnect**
  - SGI Altix
    - Intel Itanium 2
  - Cray XT3, XD1
    - AMD Opteron
- ◆ **Commodity processor with commodity interconnect**
  - Clusters
    - Pentium, Itanium, Opteron, Alpha
    - GigE, Infiniband, Myrinet, Quadrics
  - NEC TX7
  - IBM eServer
  - Dawning

Loosely  
Coupled



# Commodity Processors

## ◆ Intel Pentium Nocona

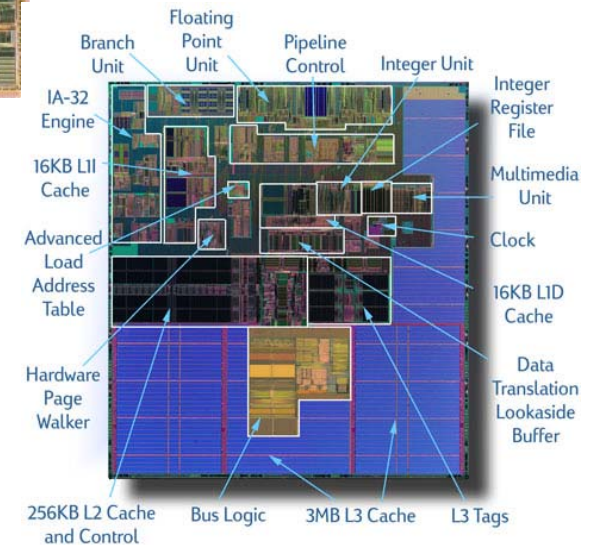
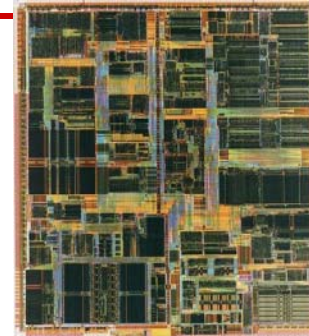
- 3.6 GHz, peak = 7.2 Gflop/s
- Linpack 100 = 1.8 Gflop/s
- Linpack 1000 = 4.2 Gflop/s

## ◆ Intel Itanium 2

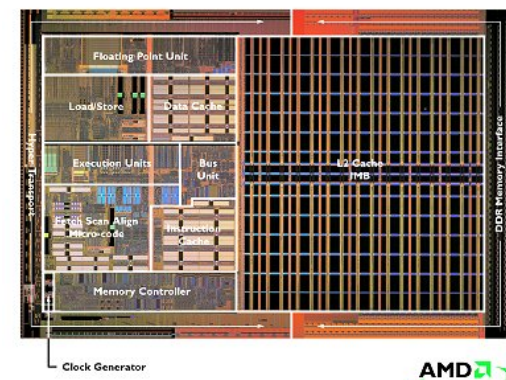
- 1.6 GHz, peak = 6.4 Gflop/s
- Linpack 100 = 1.7 Gflop/s
- Linpack 1000 = 5.7 Gflop/s

## ◆ AMD Opteron

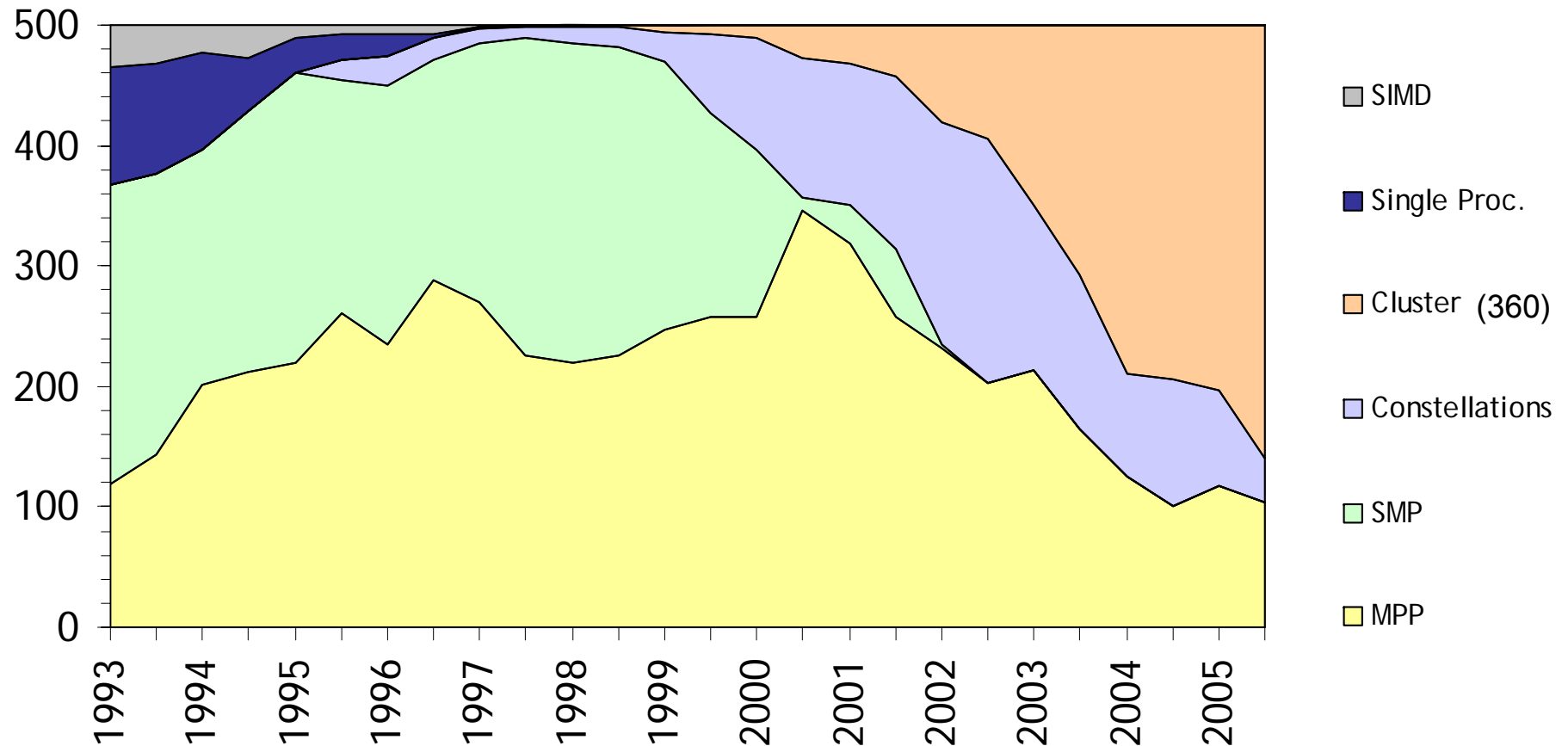
- 2.6 GHz, peak = 5.2 Gflop/s
- Linpack 100 = 1.6 Gflop/s
- Linpack 1000 = 3.9 Gflop/s



McKinley microprocessor



# Architectures / Systems



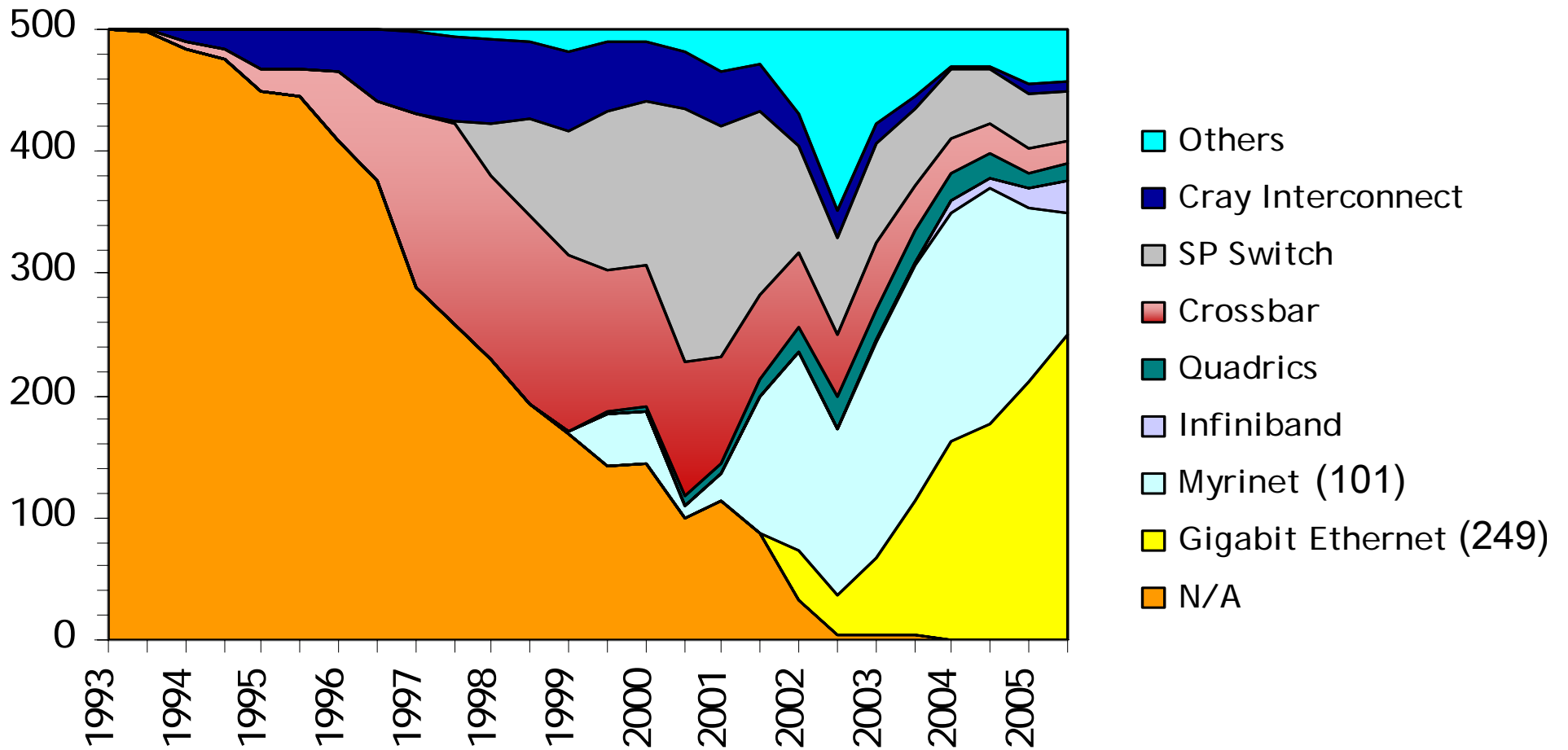
Cluster: Commodity processors & Commodity interconnect

oo

Constellation: # of procs/node  $\geq$  nodes in the system

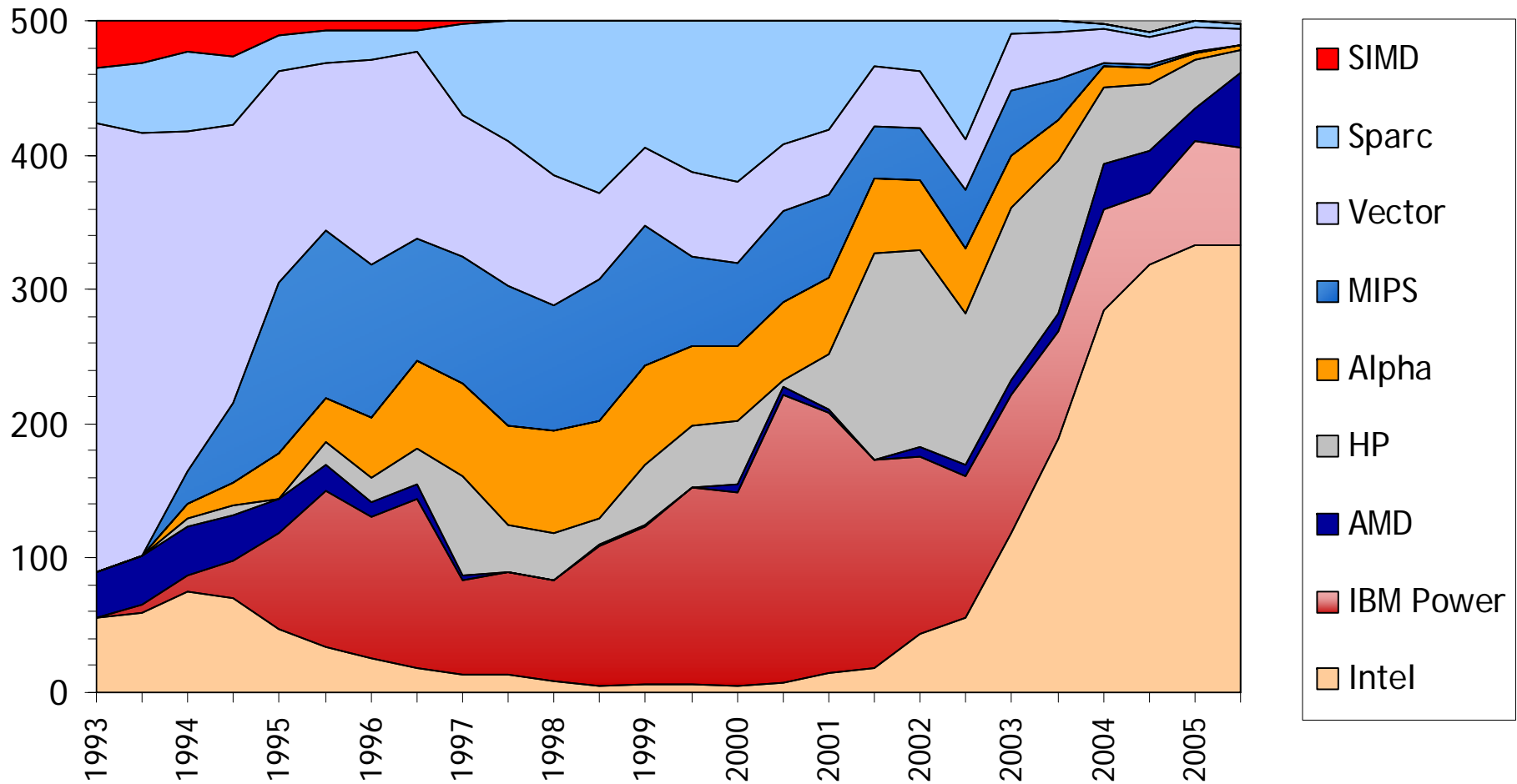


# Interconnects / Systems



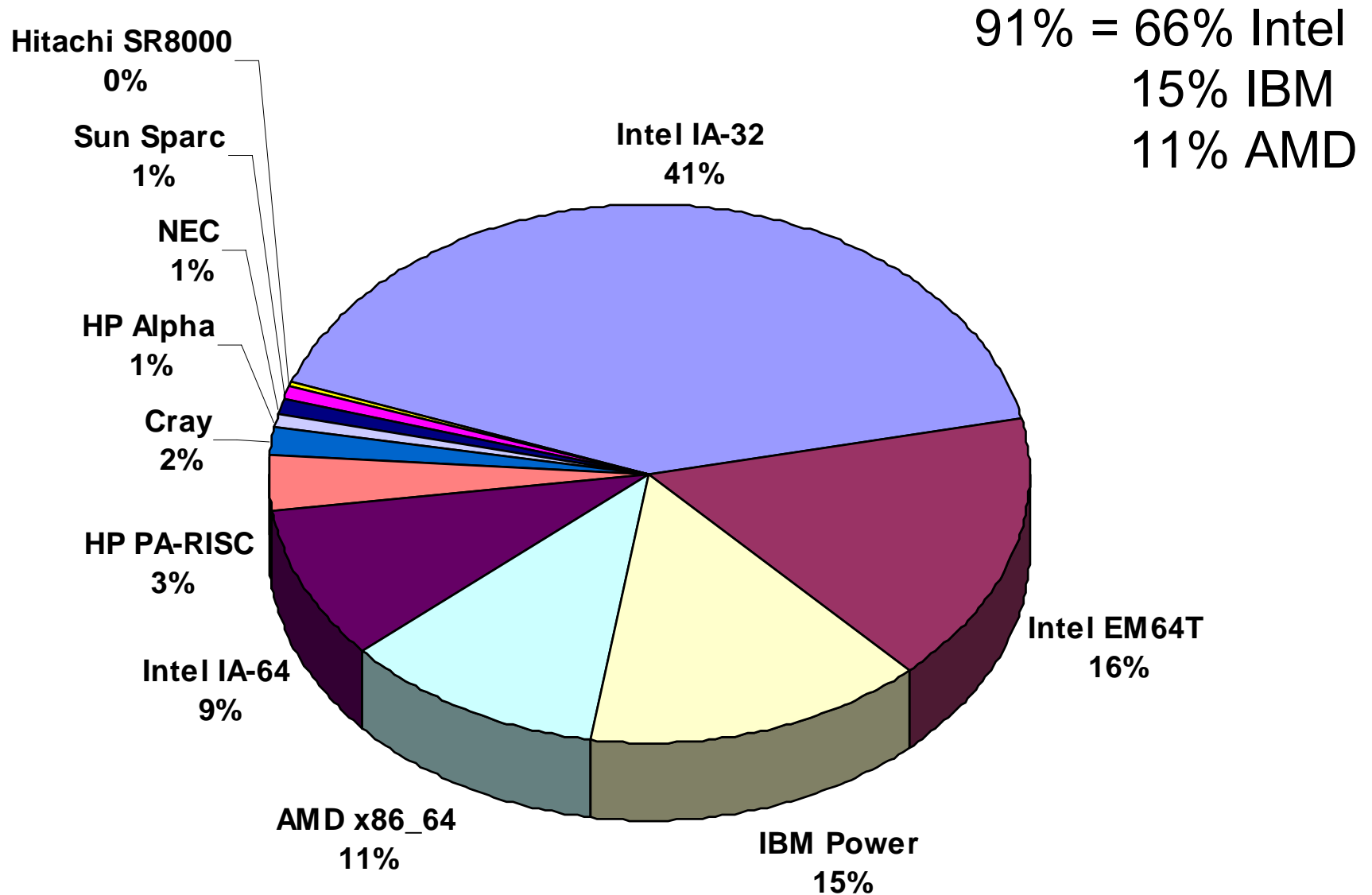


# Processor Types





# Processors Used in Each of the 500 Systems



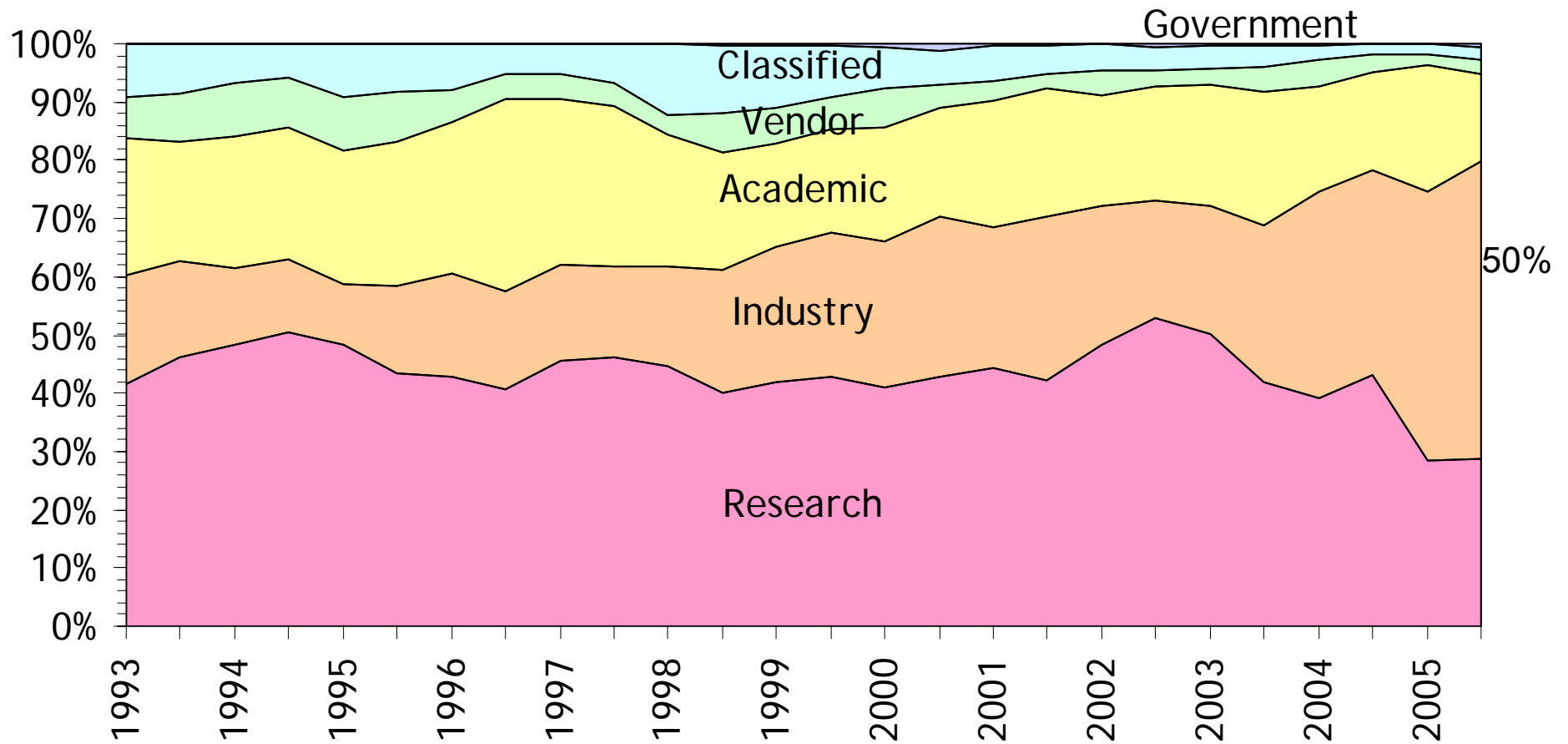


# 26th List: The TOP10

	Manufacturer	Computer	Rmax [TF/s]	Installation Site	Country	Year	#Proc
1	IBM	BlueGene/L eServer Blue Gene	280.6	DOE/NNSA/LLNL	USA	2005	131072
2	IBM	BGW eServer Blue Gene	91.29	IBM Thomas Watson	USA	2005	40960
3	IBM	ASC Purple Power5 p575	63.39	DOE/NNSA/LLNL	USA	2005	10240
<del>4</del> 3	SGI	Columbia Altix, Itanium/Infiniband	51.87	NASA Ames	USA	2004	10160
5	Dell	Thunderbird Pentium/Infiniband	38.27	Sandia	USA	2005	8000
<del>6</del> 10	Cray	Red Storm Cray XT3 AMD	36.19	Sandia	USA	2005	10880
<del>7</del> 4	NEC	Earth-Simulator SX-6	35.86	Earth Simulator Center	Japan	2002	5120
<del>8</del> 5	IBM	MareNostrum PPC 970/Myrinet	27.91	Barcelona Supercomputer Center	Spain	2005	4800
<del>9</del> 6	IBM	eServer Blue Gene	27.45	ASTRON University Groningen	Netherlands	2005	12288
10	Cray	Jaguar Cray XT3 AMD	20.53	Oak Ridge National Lab	USA	2005	5200

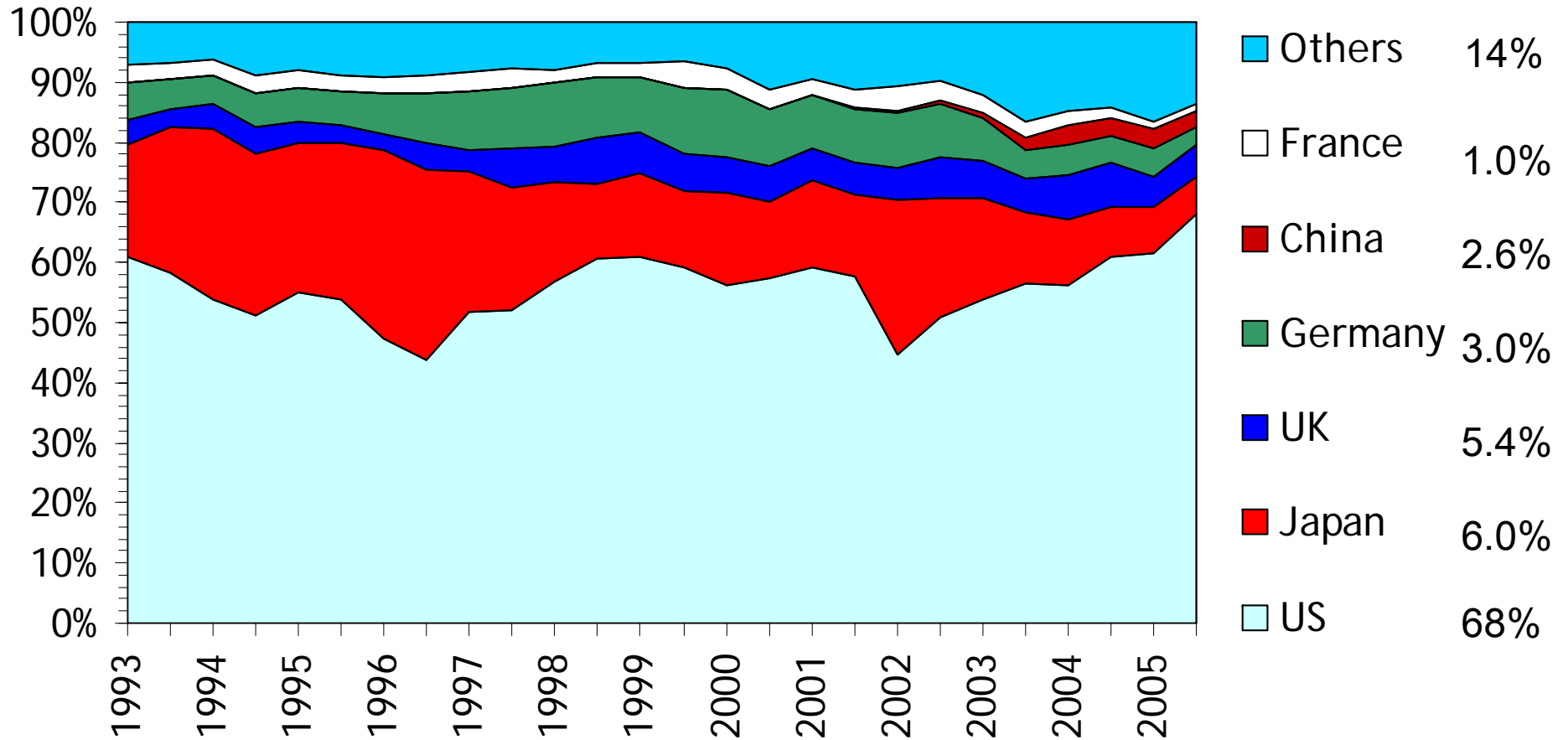


# Customer Segments / Performance



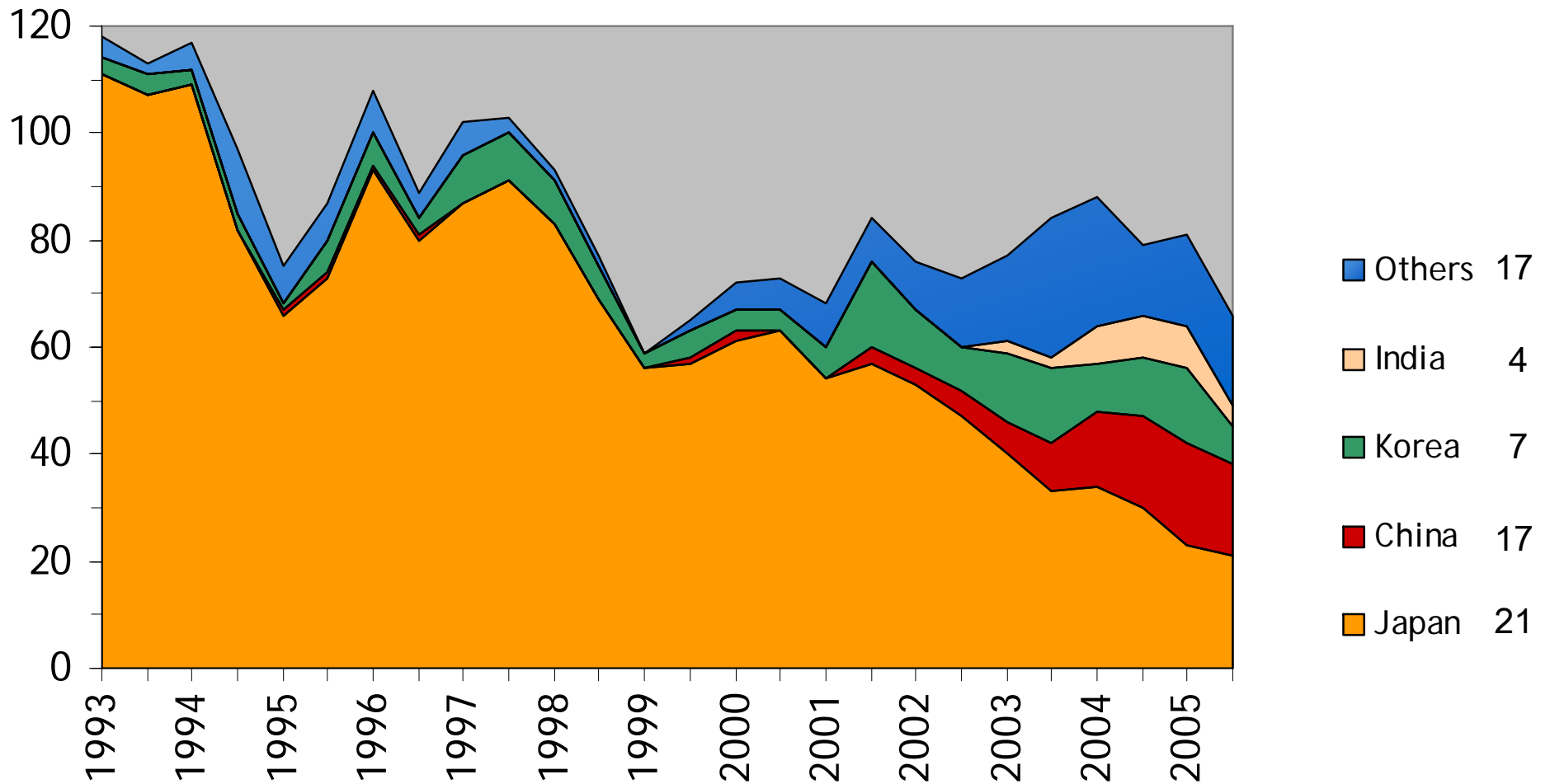


# Countries / Performance

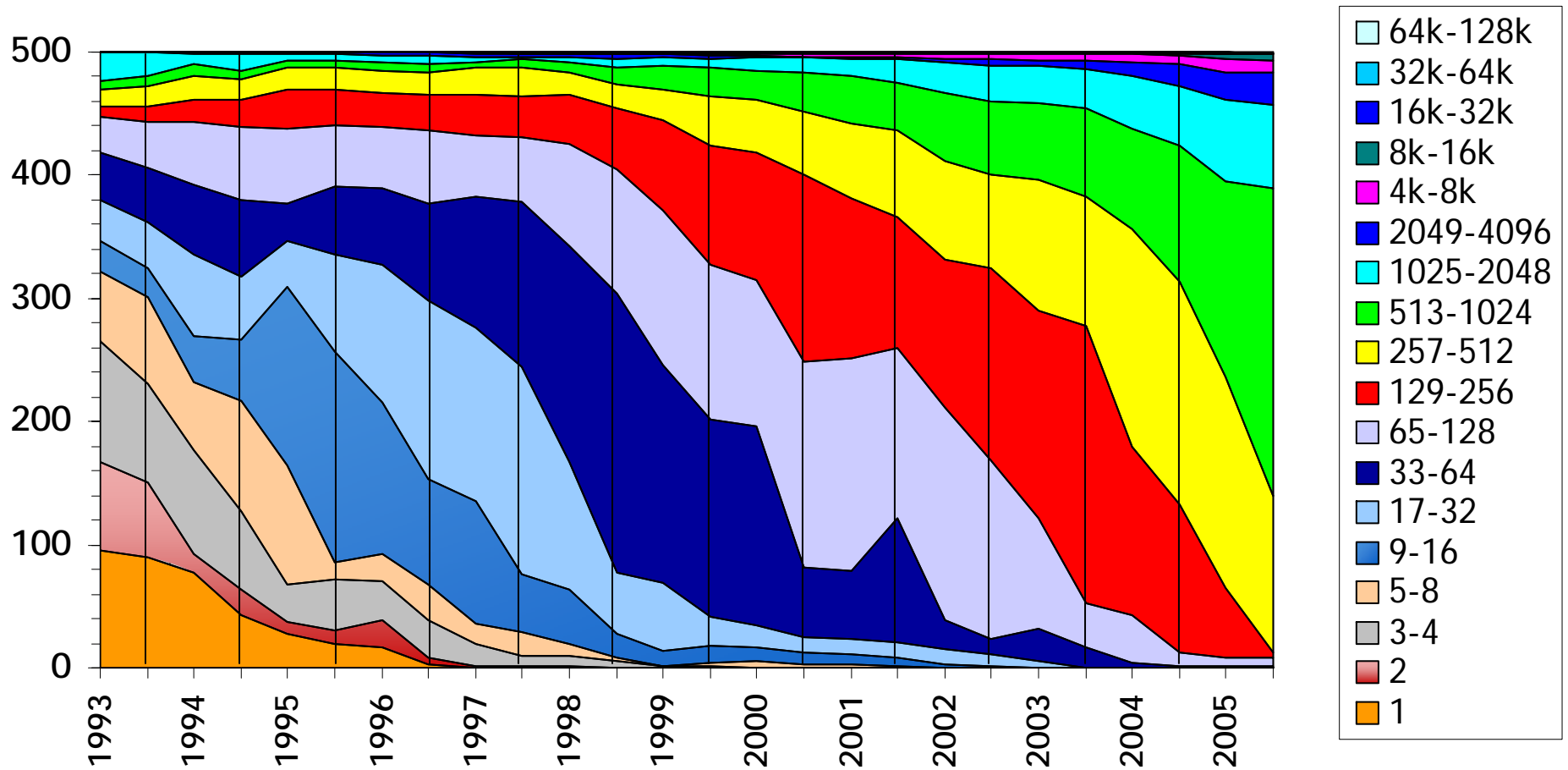




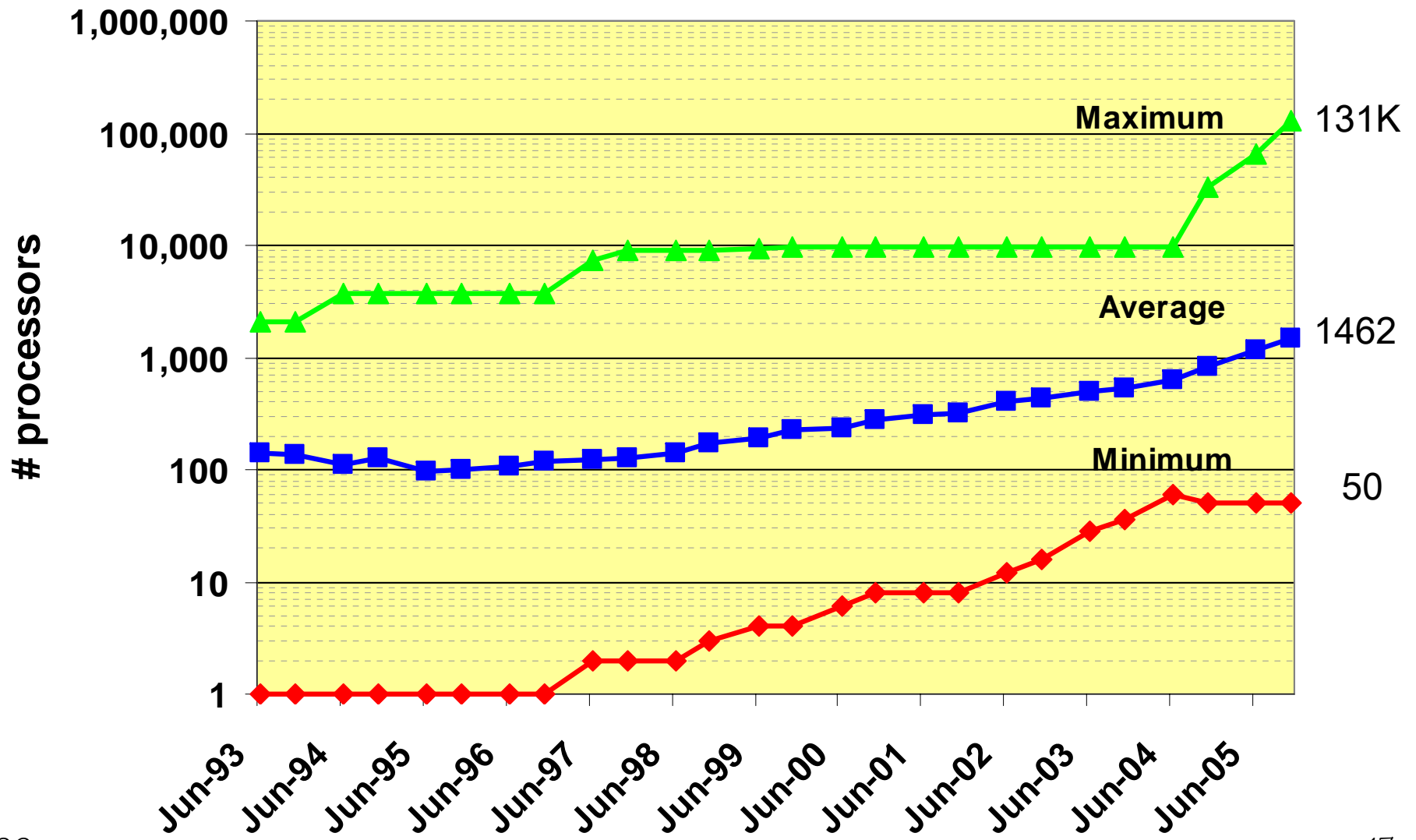
# Asian Countries / Systems

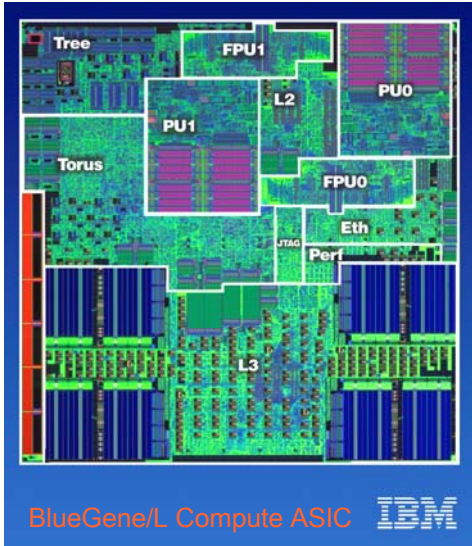


# Concurrency Levels of the Top500



# Concurrency Levels of the Top500





# IBM BlueGene/L #1 131,072 Processors

## Total of 18 systems all in the Top100

1.6 MWatts (1600 homes) (64 racks, 64x32x32)  
 43,000 ops/s/person Rack 131,072 procs

(32 Node boards, 8x8x16)  
 2048 processors

Node Board  
 (32 chips, 4x4x2)  
 16 Compute Cards  
 64 processors

Compute Card  
 (2 chips, 2x1x1)  
 4 processors

Chip  
 (2 processors)

2.8/5.6 GF/s  
 4 MB (cache)

5.6/11.2 GF/s  
 1 GB DDR

90/180 GF/s  
 16 GB DDR

2.9/5.7 TF/s  
 0.5 TB DDR

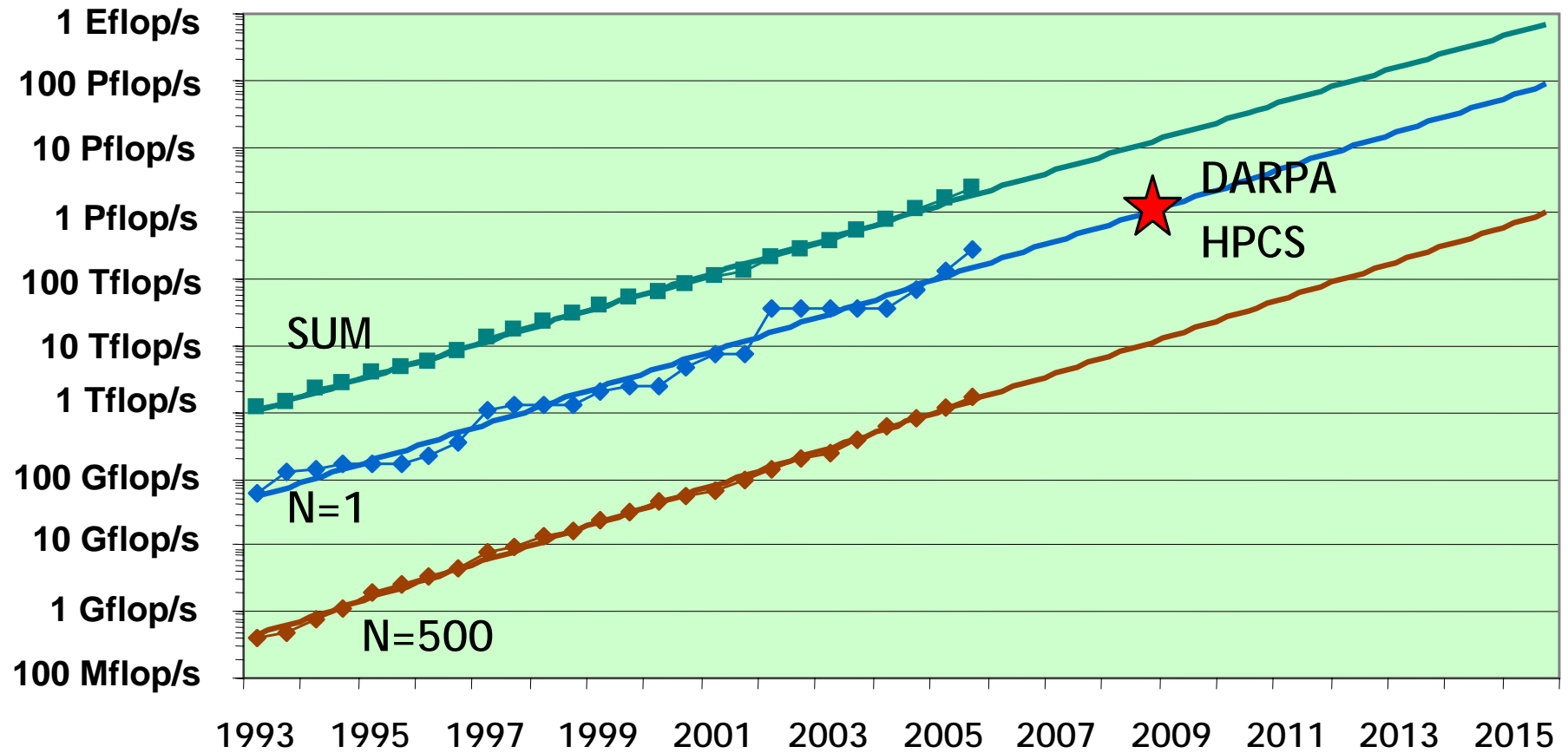
180/360 TF/s  
 32 TB DDR

Full system total of  
 131,072 processors

The compute node ASICs include all networking and processor functionality. Each compute ASIC includes two 32-bit superscalar PowerPC 440 embedded cores (note that L1 cache coherence is not maintained between these cores). (13K sec about 3.6 hours; n=1.8M)

**"Fastest Computer"**  
 BG/L 700 MHz 131K proc  
 64 racks  
 Peak: 367 Tflop/s  
 Linpack: 281 Tflop/s  
 77% of peak

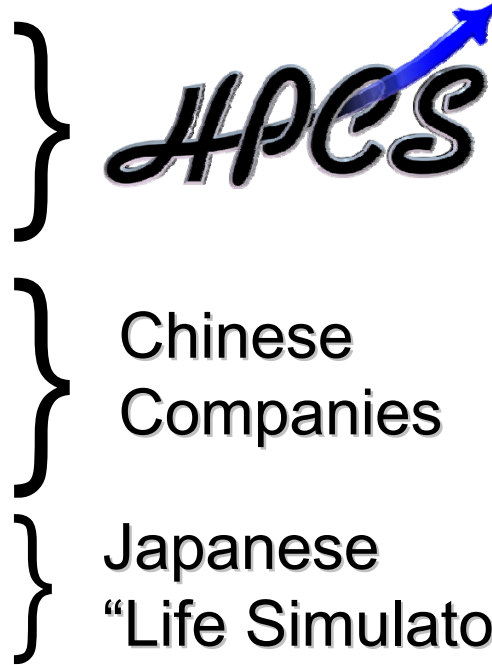
# Performance Projection



# A PetaFlop Computer by the End of the Decade

- ◆ 10 Companies working on a building a Petaflop system by the end of the decade.

- Cray
- IBM
- Sun
- Dawning
- Galactic
- Lenovo
- Hitachi
- NEC
- Fujitsu
- Bull



Chinese Companies

Japanese

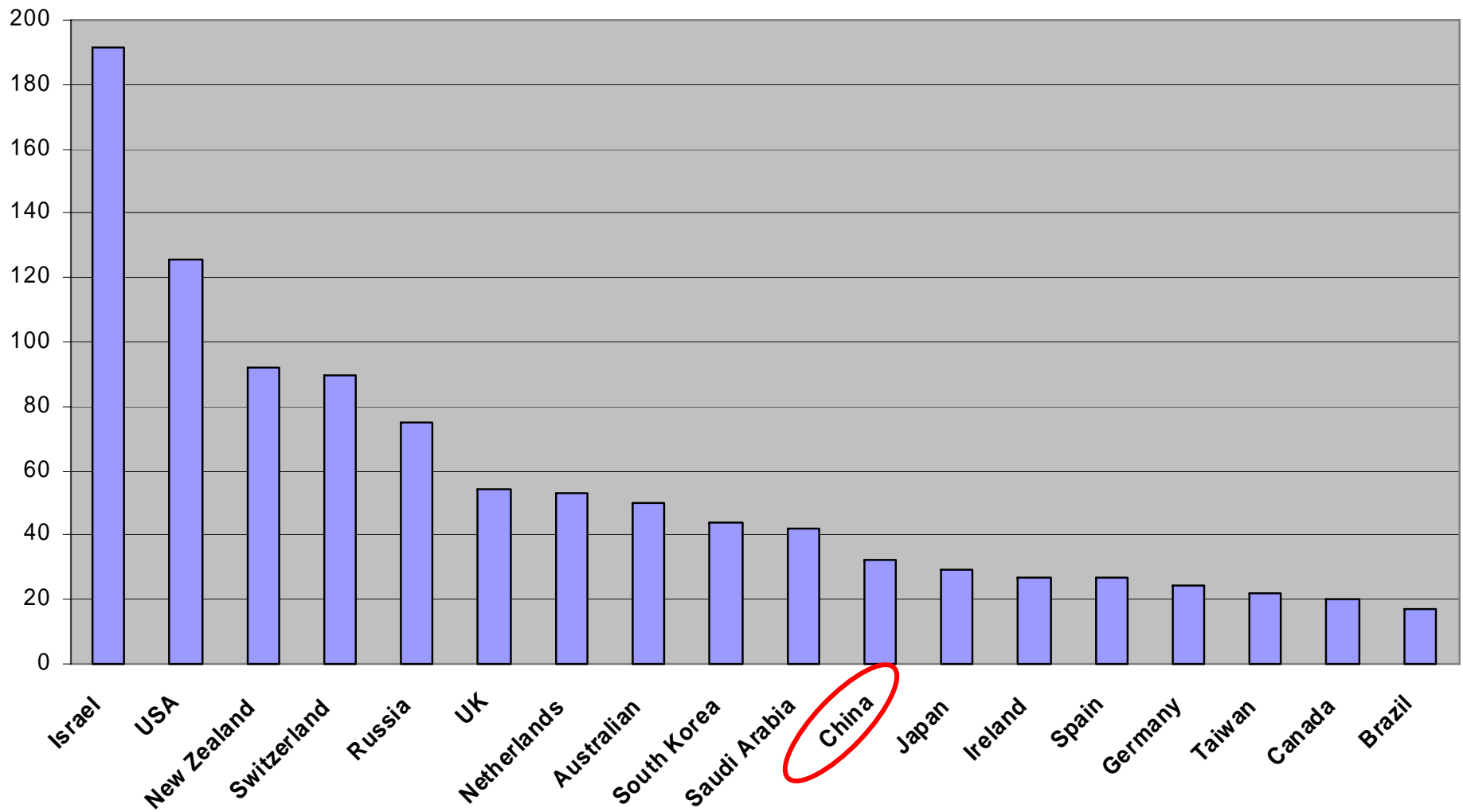
“Life Simulator” (10 Pflop/s)





# Flops per Gross Domestic Product

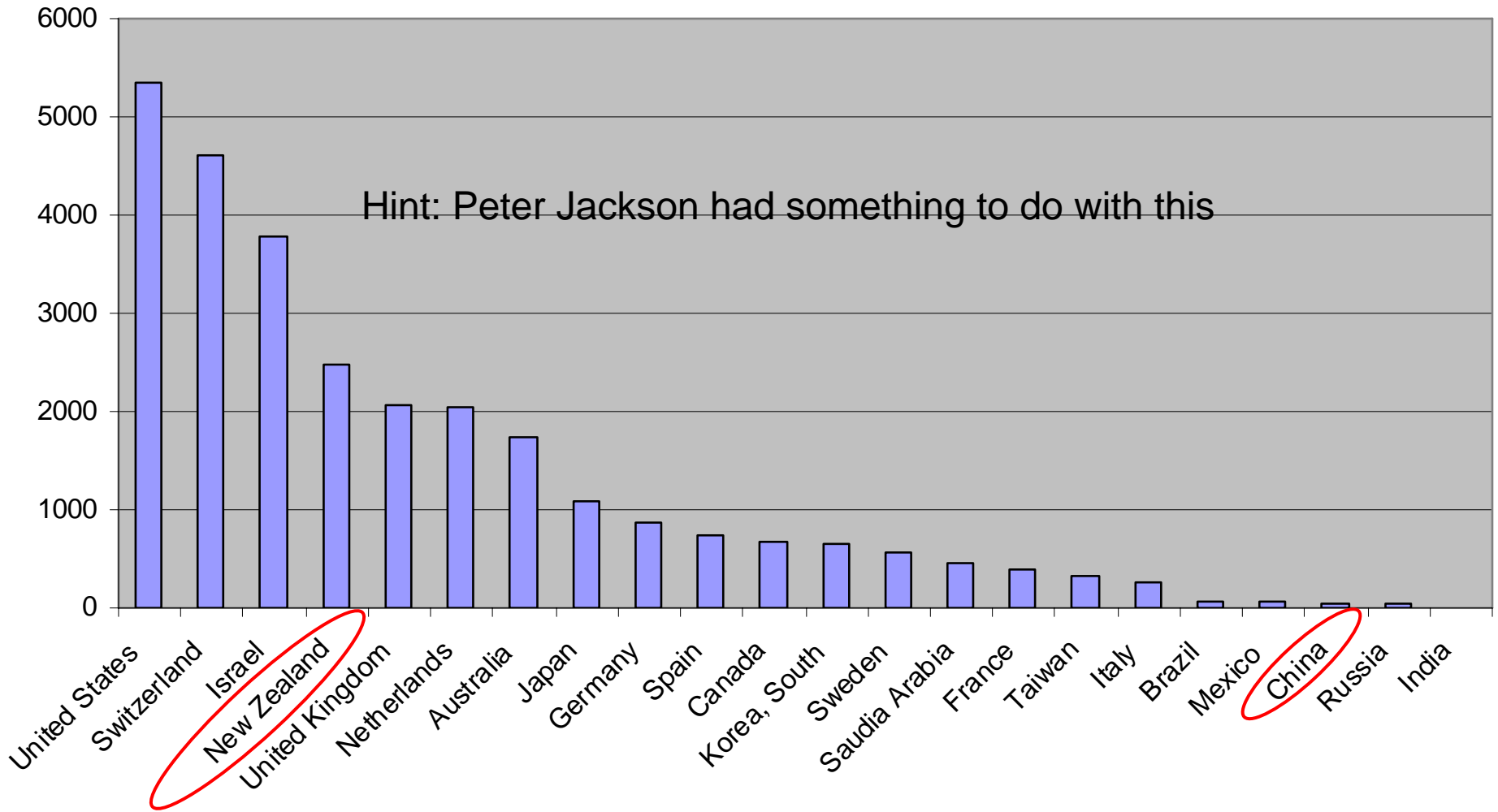
Based on the November 2005 Top500





# KFlop/s per Capita (Flops/Pop)

Based on the November 2005 Top500



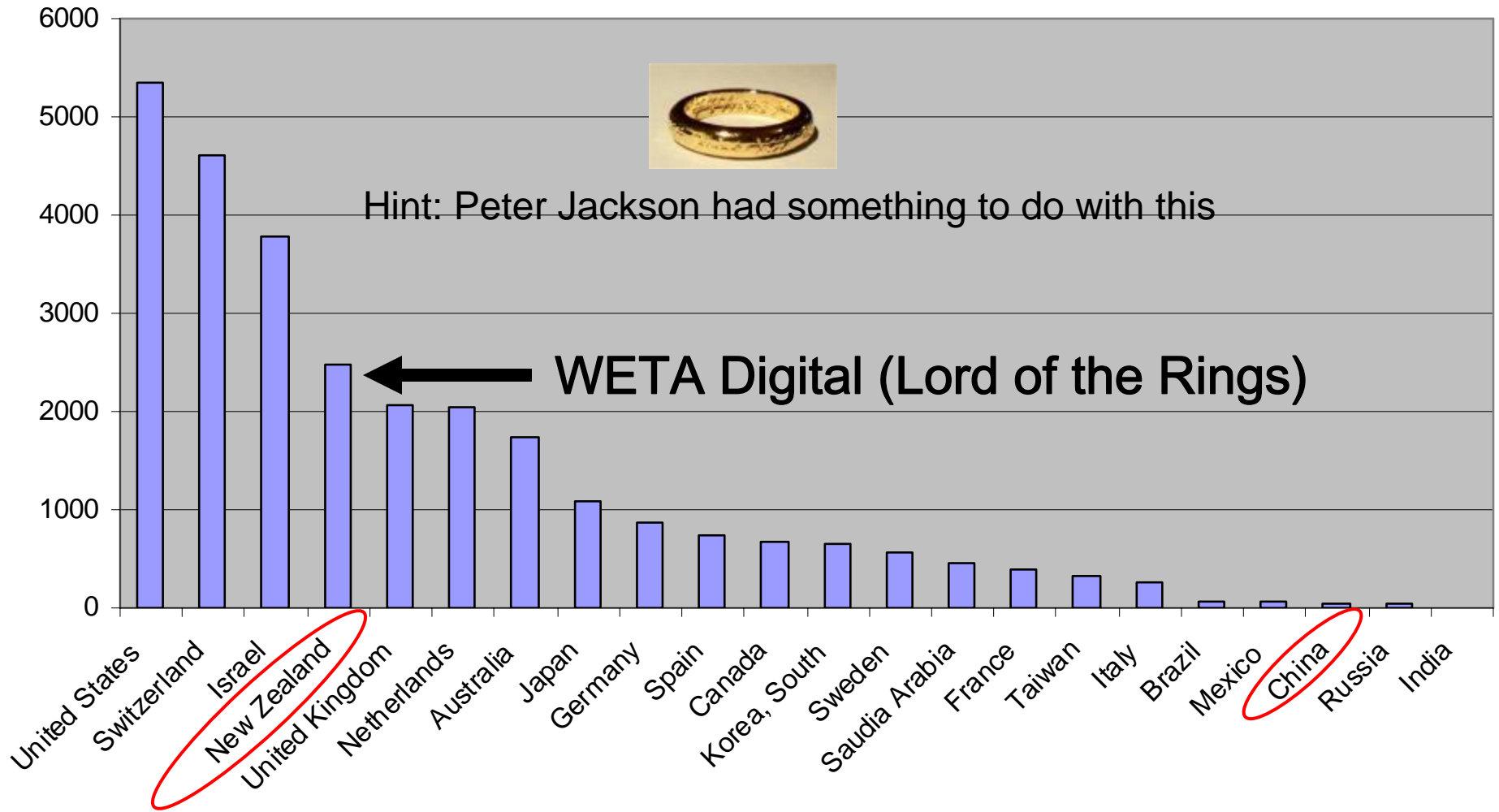
00

Has nothing to do with the 47.2 million sheep in NZ

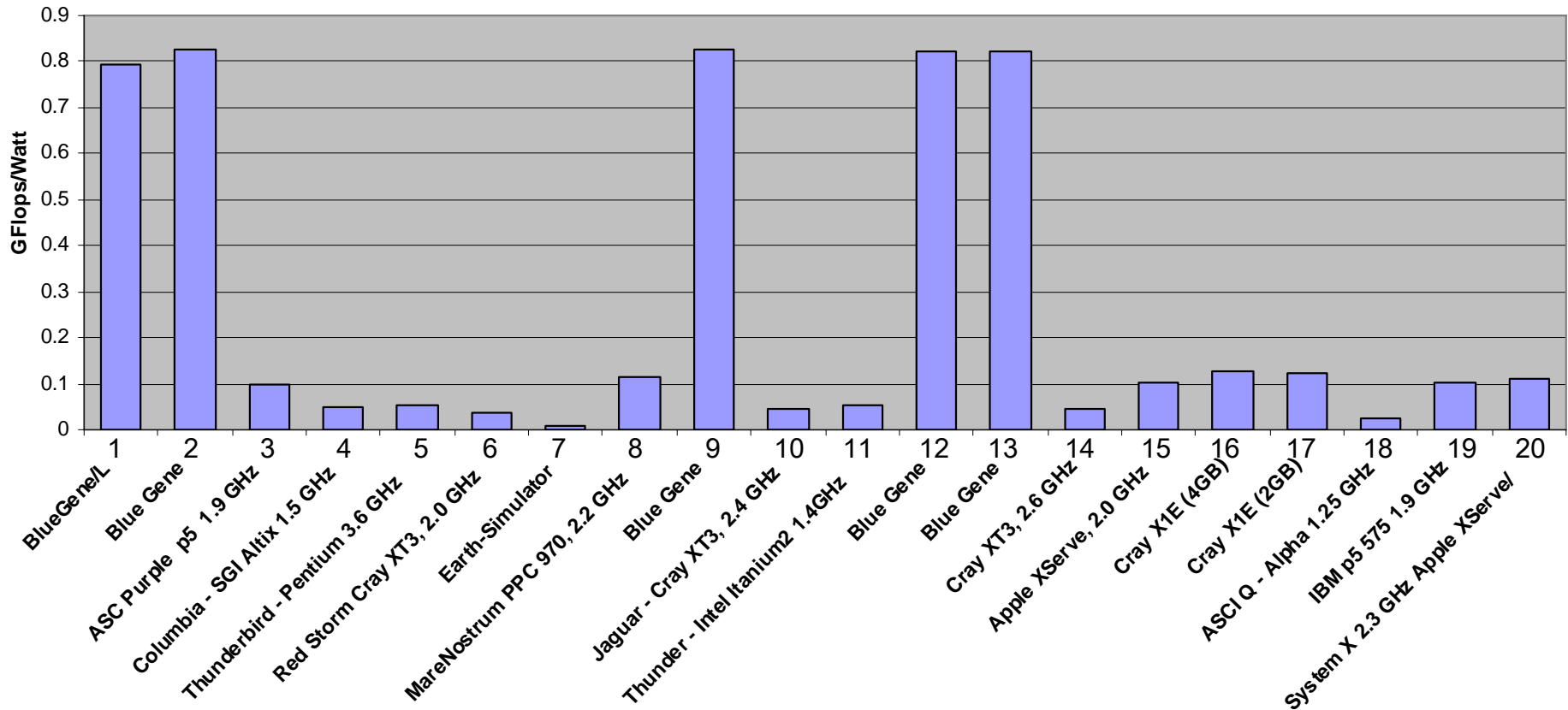


# KFlop/s per Capita (Flops/Pop)

Based on the November 2005 Top500

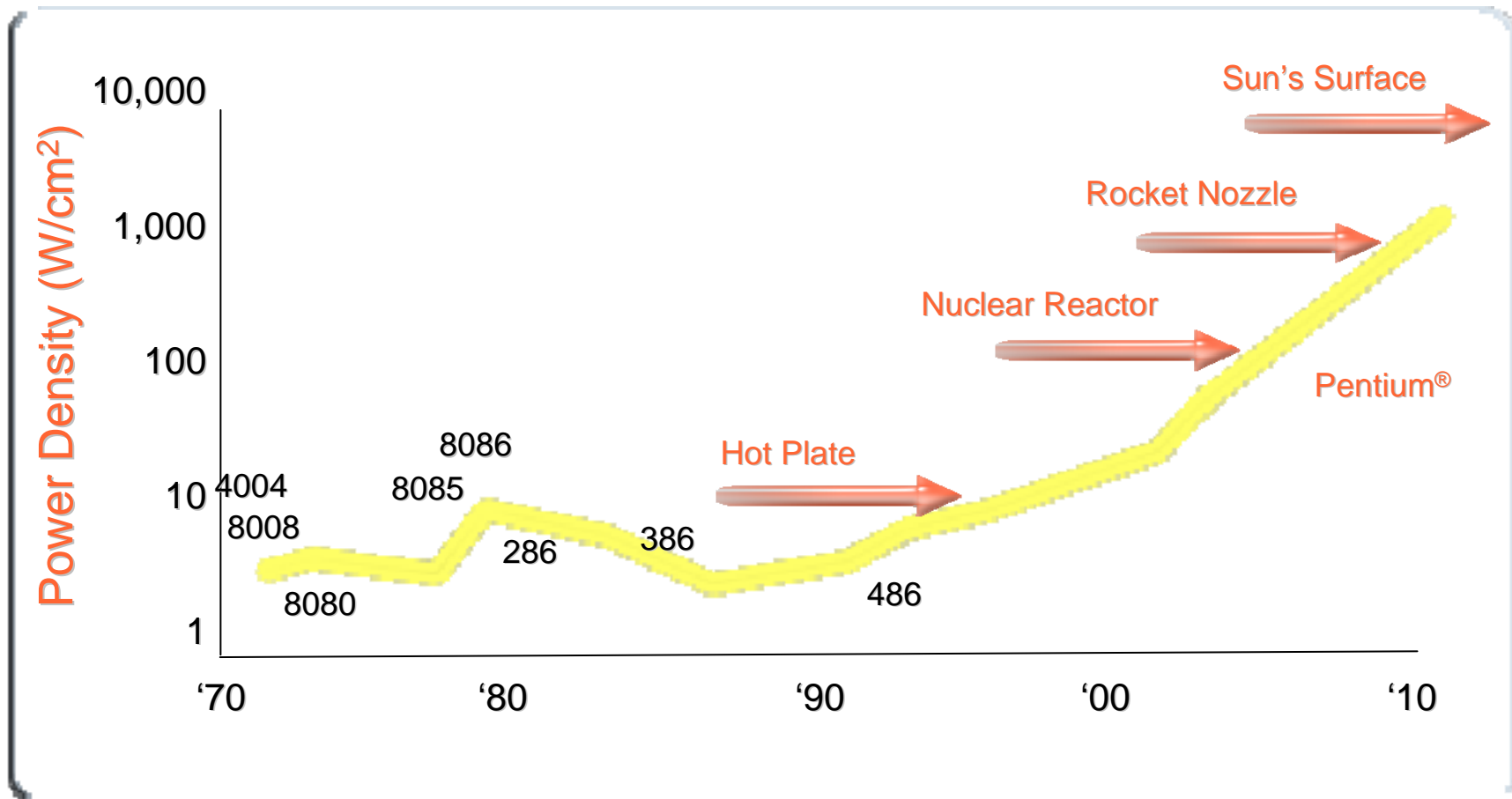


# Fuel Efficiency: GFlops/Watt



# Today's CPU Architecture:

## Heat becoming an unmanageable problem

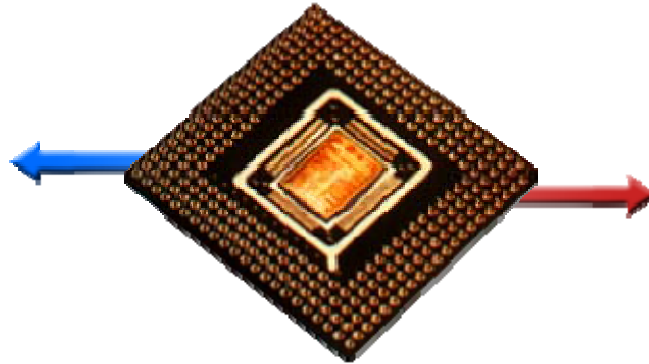


Intel Developer Forum, Spring 2004 - Pat Gelsinger  
(Pentium at 90 W)

Cube relationship between the cycle time and power. 2.5

# Increasing CPU Performance: A Delicate Balancing Act

Lower Voltage



Increase Clock Rate & Transistor Density

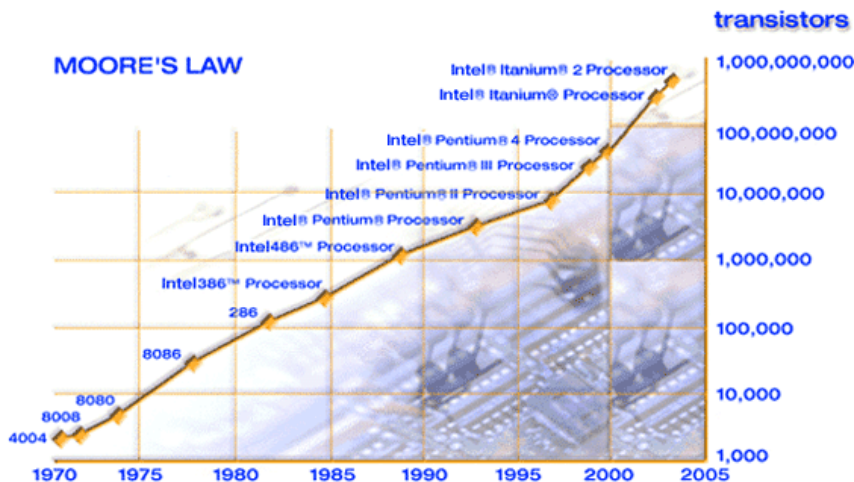
We have seen increasing number of gates on a chip and increasing clock speed.

Heat becoming an unmanageable problem, Intel Processors > 100 Watts

We will not see the dramatic increases in clock speeds in the future.

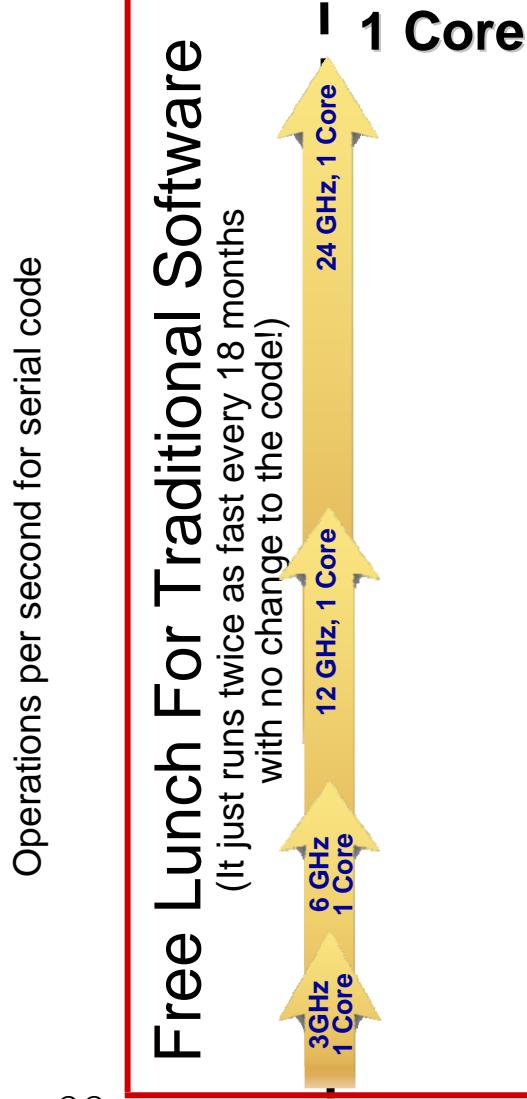
However, the number of gates on a chip will continue to increase.

Intel Yonah doubles the processing power on a per watt basis. <sup>26</sup>





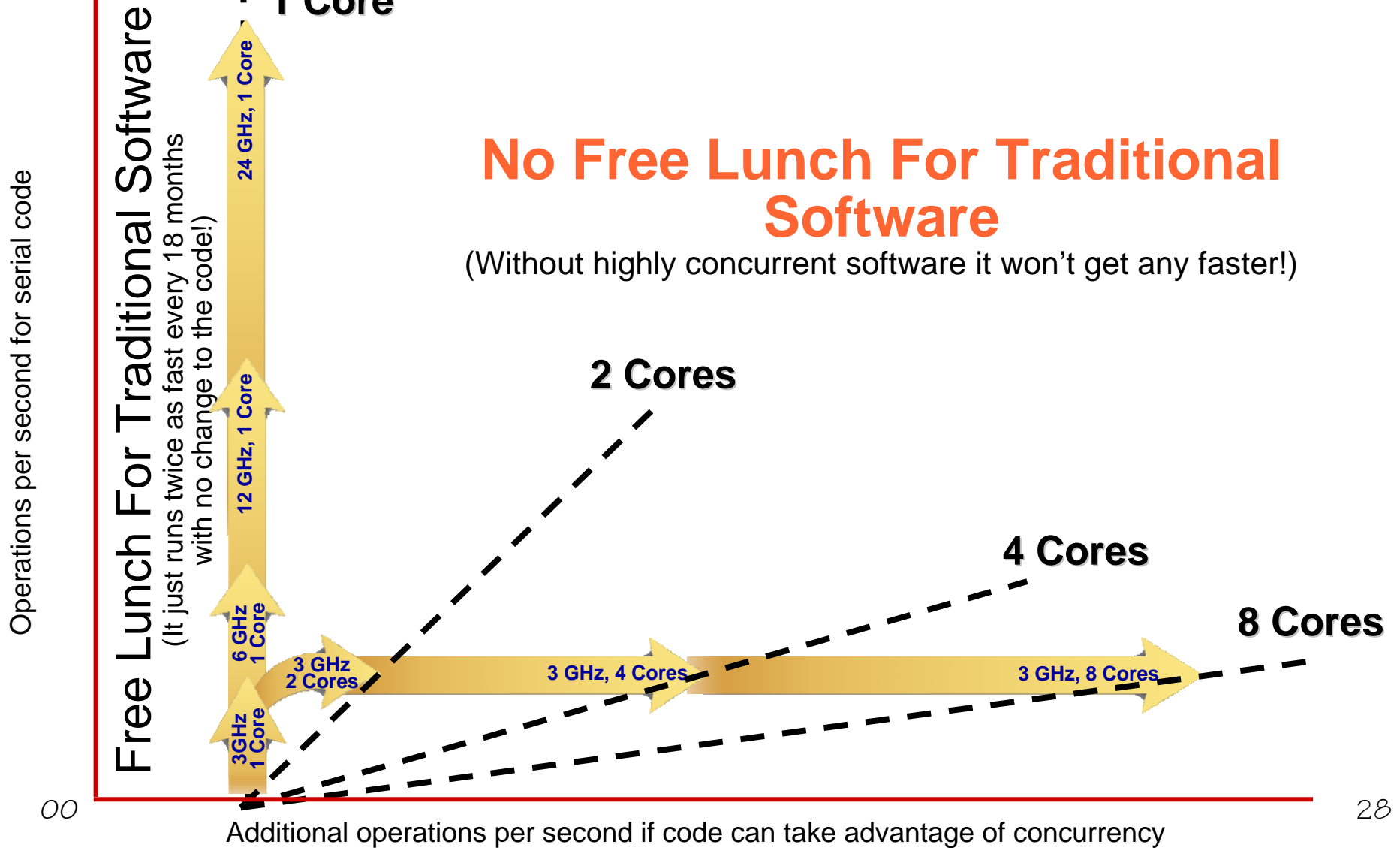
# Change Is Coming



Additional operations per second if code can take advantage of concurrency

From Craig Mundie, Microsoft

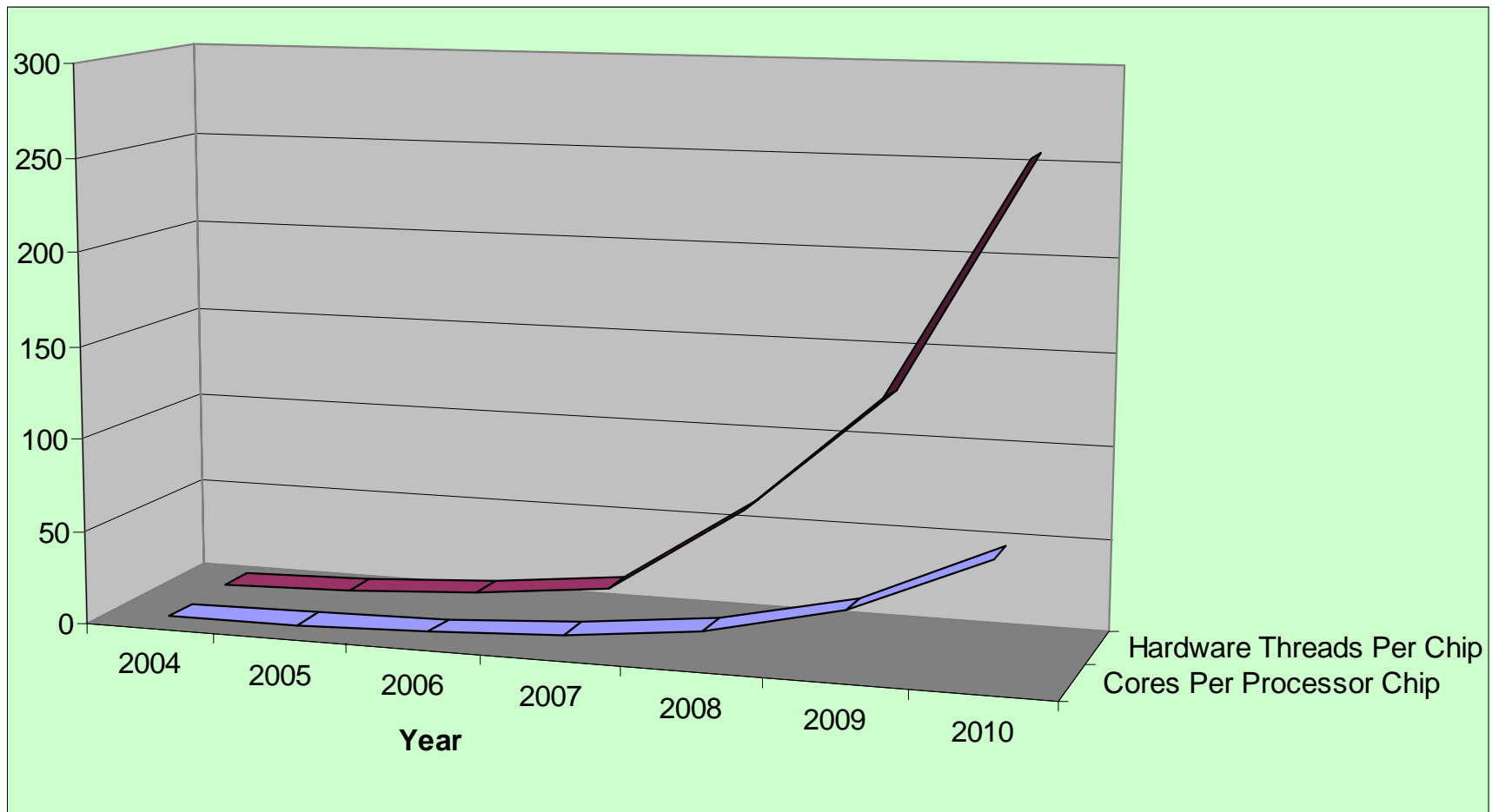
# Change Is Coming





# CPU Desktop Trends 2004-2010

- ◆ Relative processing power will continue to double every 18 months
- ◆ 256 logical processors per chip in late 2010





# Commodity Processor Trends

## Bandwidth/Latency is the Critical Issue, not FLOPS



Got Bandwidth?

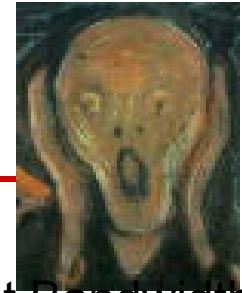
	Annual increase	Typical value in 2005
Single-chip floating-point performance	59%	4 GFLOP/s
Front-side bus bandwidth	23%	1 GWord/s = 0.25 word/flop
DRAM latency	(5.5%)	70 ns = 280 FP ops = 70 loads

00 Source: *Getting Up to Speed: The Future of Supercomputing*, National Research Council, 222 pages, 2004, National Academies Press, Washington DC, ISBN 0-309-09502-6.



# Commodity Processor Trends

Bandwidth/Latency is the Critical Issue, not FLOPS



Got Bandwidth?

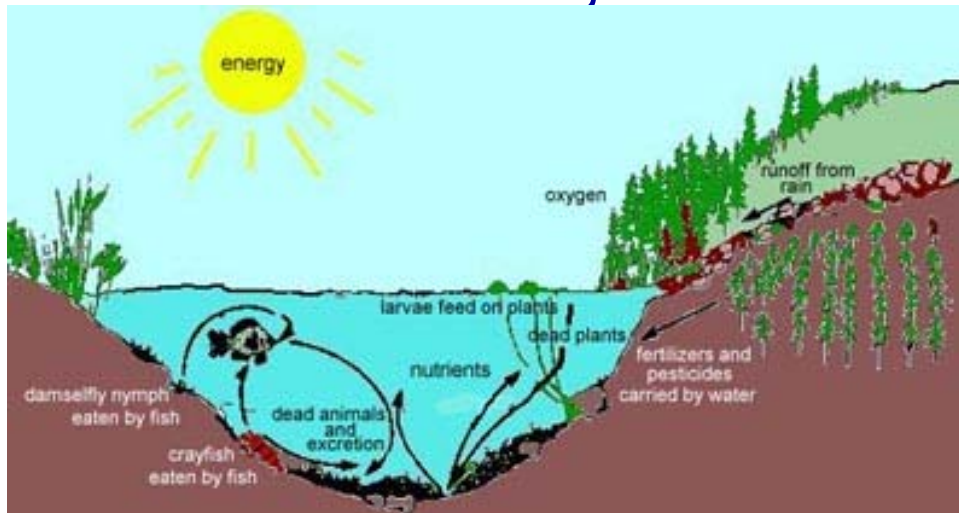
	Annual increase	Typical value in 2005	Typical value in 2010	Typical value in 2020
Single-chip floating-point performance	59%	4 GFLOP/s	32 GFLOP/s	3300 GFLOP/s
Front-side bus bandwidth	23%	1 GWord/s = 0.25 word/flop	3.5 GWord/s = 0.11 word/flop	27 GWord/s = 0.008 word/flop
DRAM latency	(5.5%)	70 ns = 280 FP ops = 70 loads	50 ns = 1600 FP ops = 170 loads	28 ns = 94,000 FP ops = 780 loads

00 Source: *Getting Up to Speed: The Future of Supercomputing*, National Research Council, 222 pages, 2004, National Academies Press, Washington DC, ISBN 0-309-09502-6.

# Future Challenge: Developing the Ecosystem for HPC

From the NRC Report on "The Future of Supercomputing":

- ◆ Hardware, software, algorithms, tools, networks, institutions, applications, and people who solve supercomputing applications can be thought of collectively as a multifaceted ecosystem
- ◆ Research investment in HPC should be informed by the ecosystem point of view - progress must come on a broad front of interrelated technologies, rather than in the form of individual breakthroughs.



A supercomputer ecosystem is a continuum of computing platforms, system software, algorithms, tools, networks, and the people who know how to exploit them to solve computational science applications.

# Real Crisis With HPC Is With The Software

---

- ◆ Our ability to configure a hardware system capable of 1 PetaFlop ( $10^{15}$  ops/s) is without question just a matter of time and \$\$.
- ◆ A supercomputer application and software are usually much more long-lived than a hardware
  - Hardware life typically five years at most... Apps 20-30 years
  - Fortran and C are the main programming models (still!!!)
- ◆ The REAL CHALLENGE is Software
  - Programming hasn't changed since the 70's
  - HUGE manpower investment
    - MPI... is that all there is?
  - Often requires HERO programming
  - Investments in the entire software stack is required (OS, libs, etc.)
- ◆ Software is a major cost component of modern technologies.
  - The tradition in HPC system procurement is to assume that the software is free... SOFTWARE COSTS (over and over)

# Summary of Current Unmet Needs

---

- ◆ Performance / Portability
- ◆ Fault tolerance
- ◆ Memory bandwidth/Latency
- ◆ Adaptability: Some degree of autonomy to self optimize, test, or monitor.
  - Able to change mode of operation: static or dynamic
- ◆ Better programming models
  - Global shared address space
  - Visible locality
- ◆ Maybe coming soon (incremental, yet offering real benefits):
  - Global Address Space (GAS) languages: UPC, Co-Array Fortran, Titanium, Chapel
    - "Minor" extensions to existing languages
    - More convenient than MPI
    - Have performance transparency via explicit remote memory references
- ◆ What's needed is a long-term, balanced investment in hardware, software, algorithms and applications.

# Collaborators / Support

<http://www.top500.org/>

## ◆ Top500 Team

- Erich Strohmaier, NERSC
- Hans Meuer, Mannheim
- Horst Simon, NERSC



[Advertising Programs](#) - [Business Solutions](#) - [About Google](#)

[Make Google Your Homepage!](#)

©2005 Google

- ◆ Slides are available at my website.