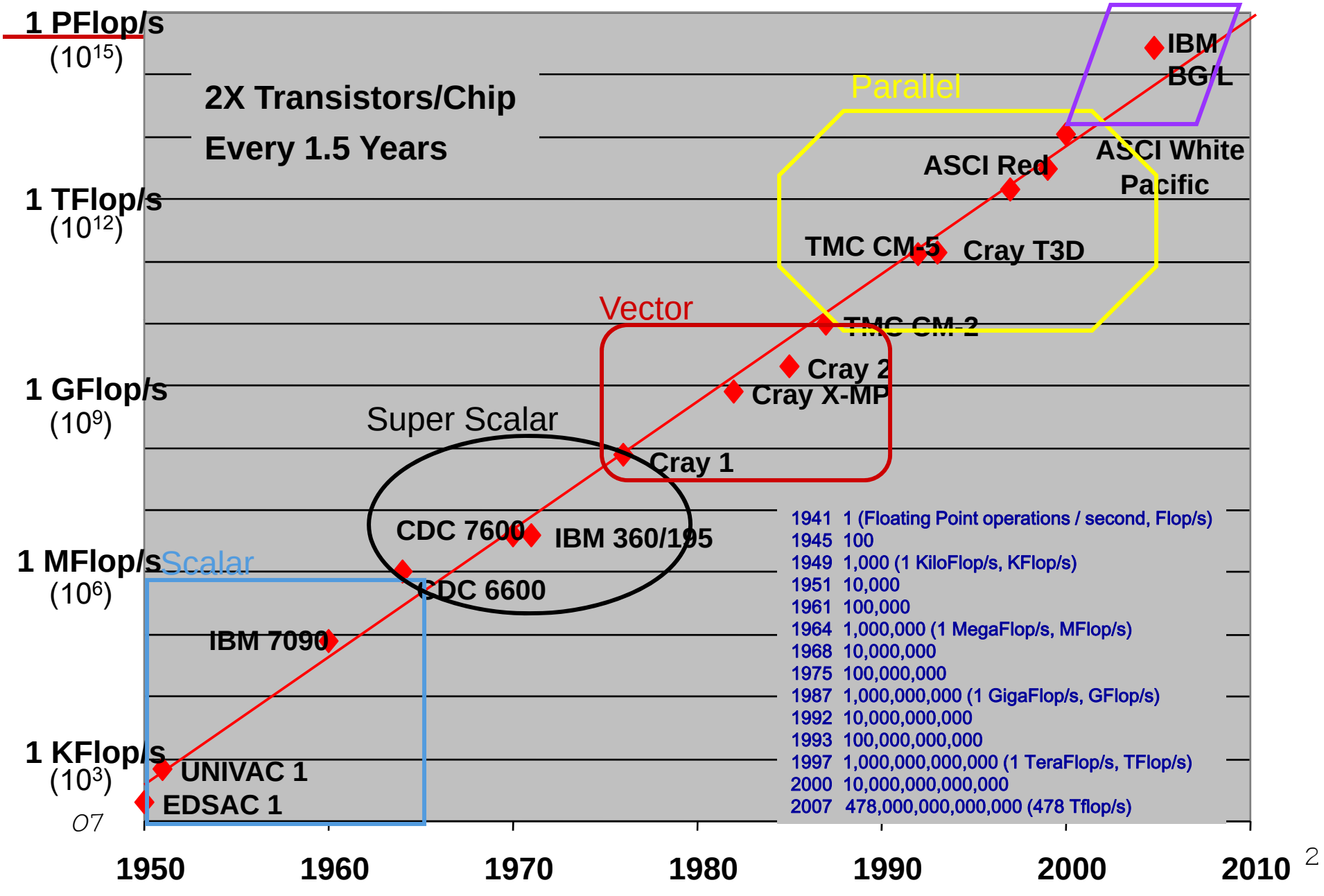
November 2007 Top500

Jack Dongarra
University of Tennessee
and
Oak Ridge National Laboratory

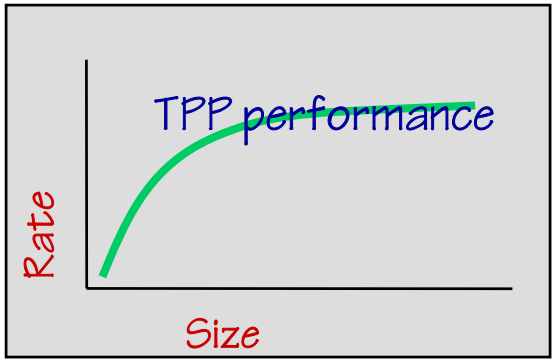
A Growth-Factor of a Billion in Performance in a Career



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

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

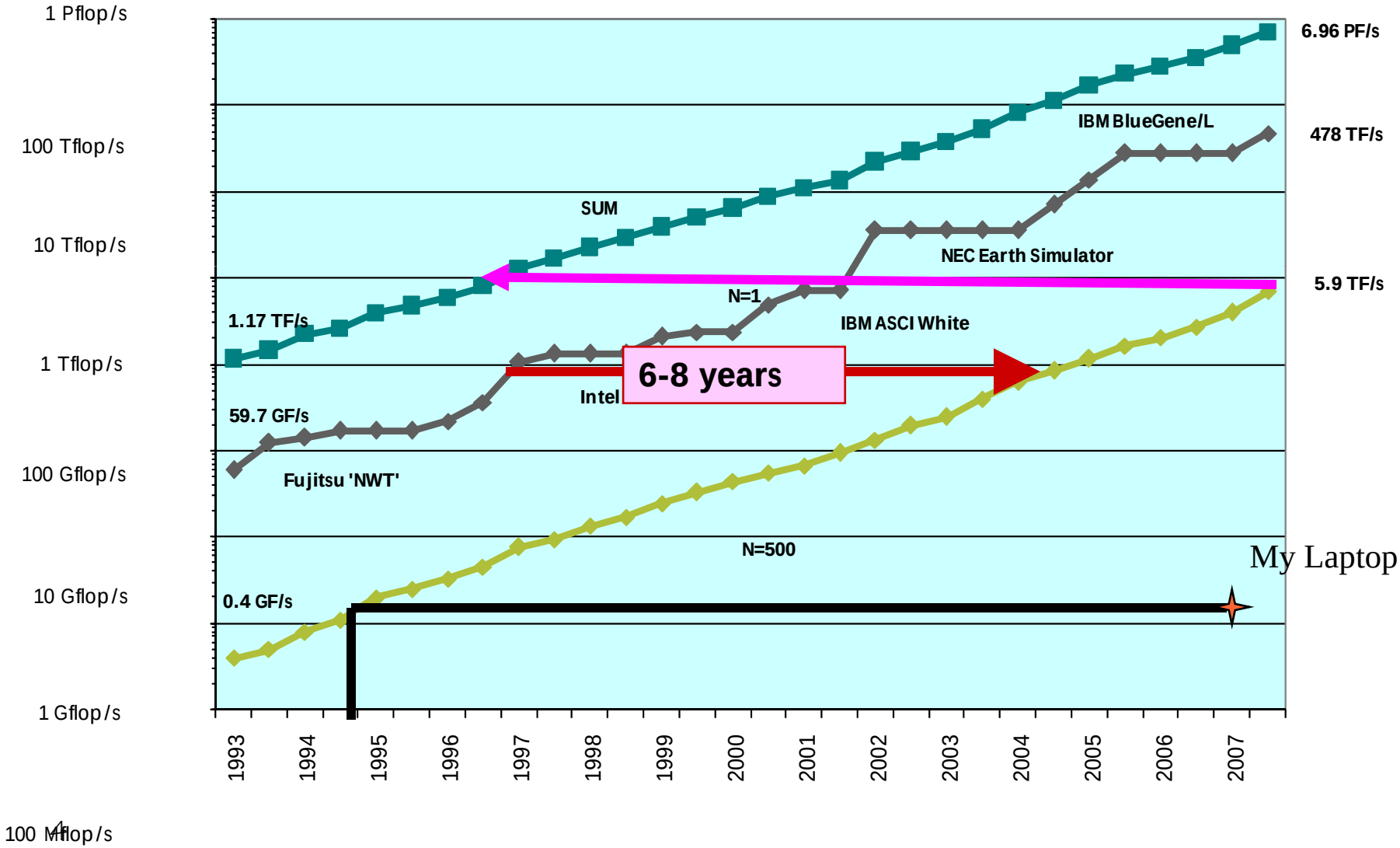
$Ax=b$, *dense problem*



- Updated twice a year
  SC‘xy in the States in November
  Meeting in Germany in June

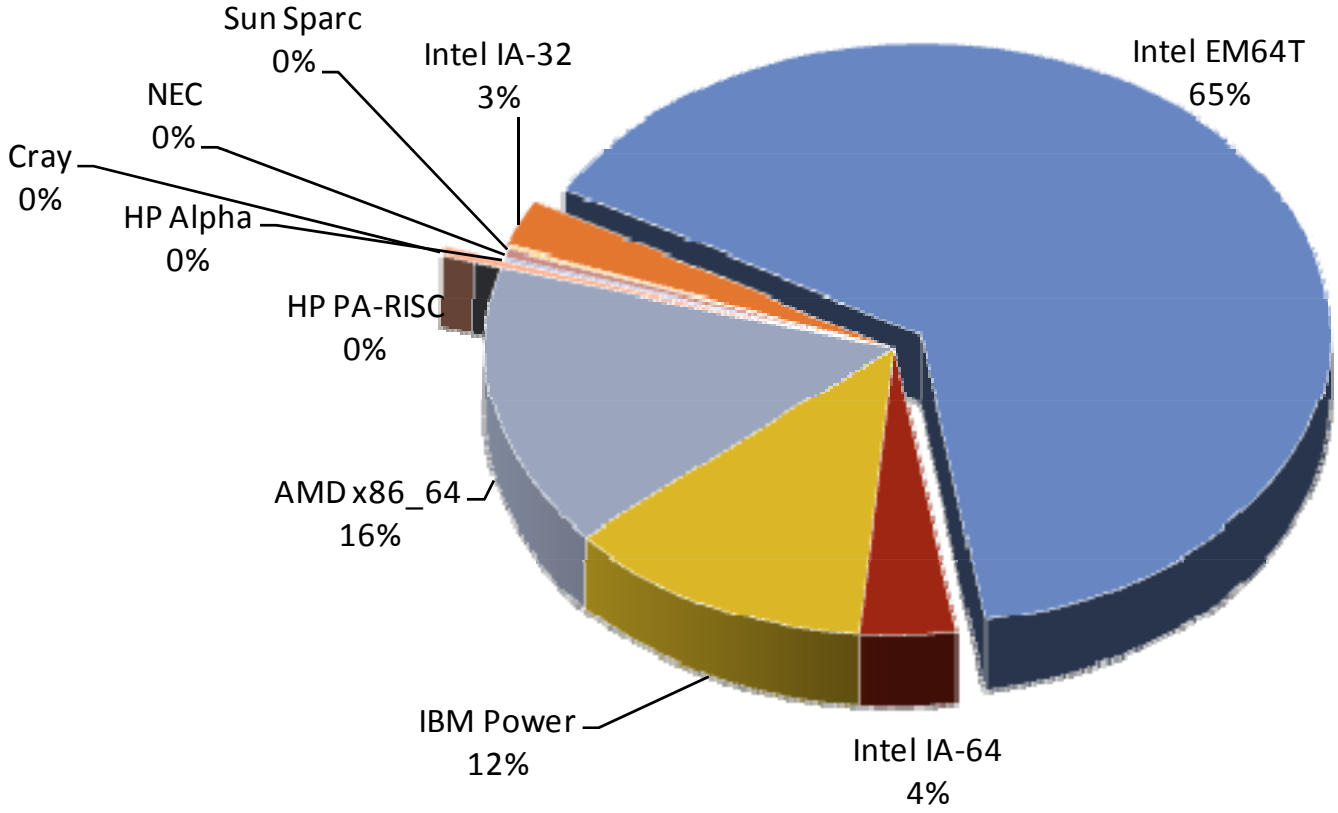
- All data available from **www.top500.org**

Performance Development

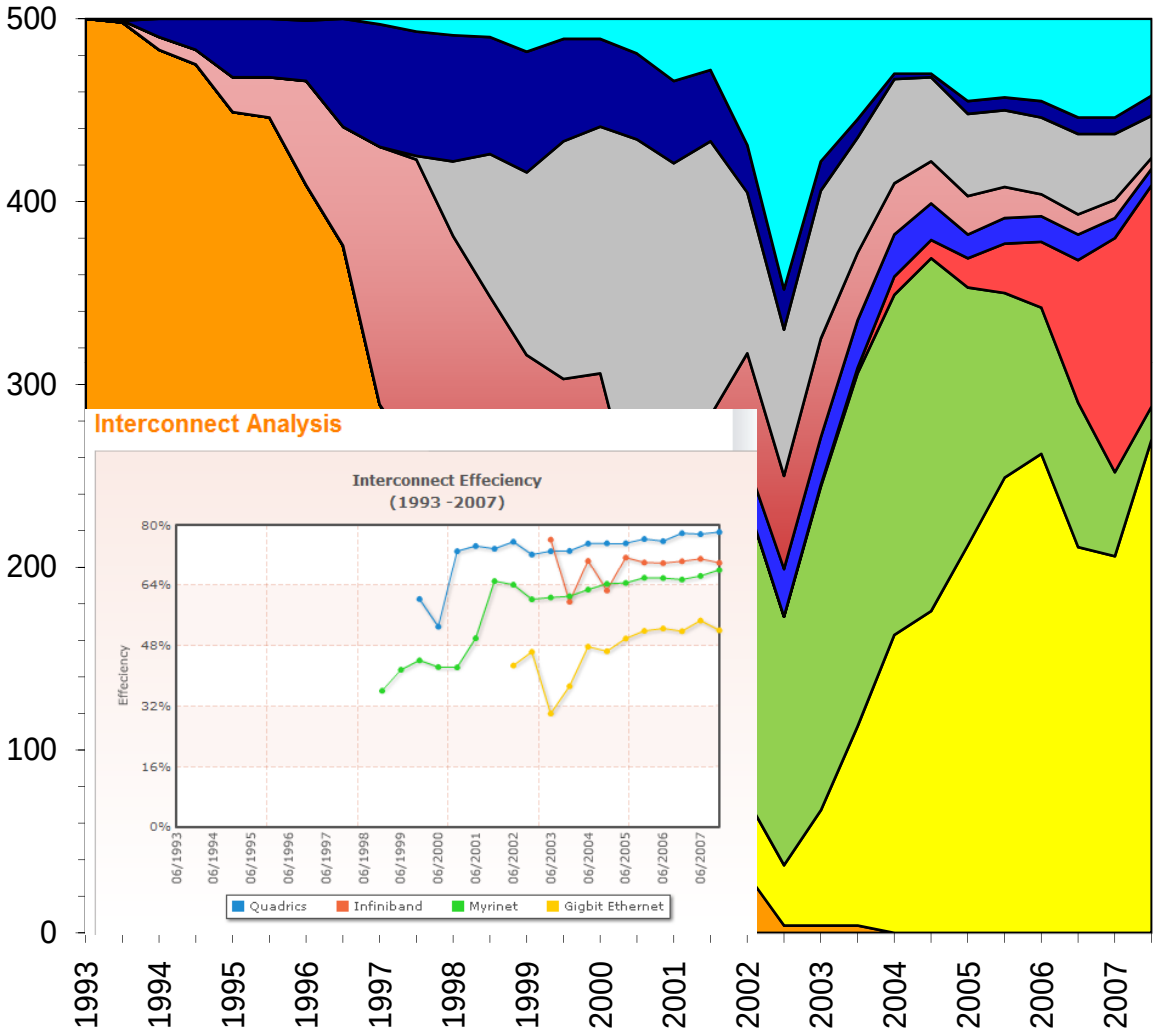


Chips Used in Each of the 500 Systems

72% Intel
12% IBM
16% AMD



Interconnects / Systems



Others

Cray Interconnect

SP Switch

Crossbar

Quadrics

Infiniband (121)

Myrinet (18)

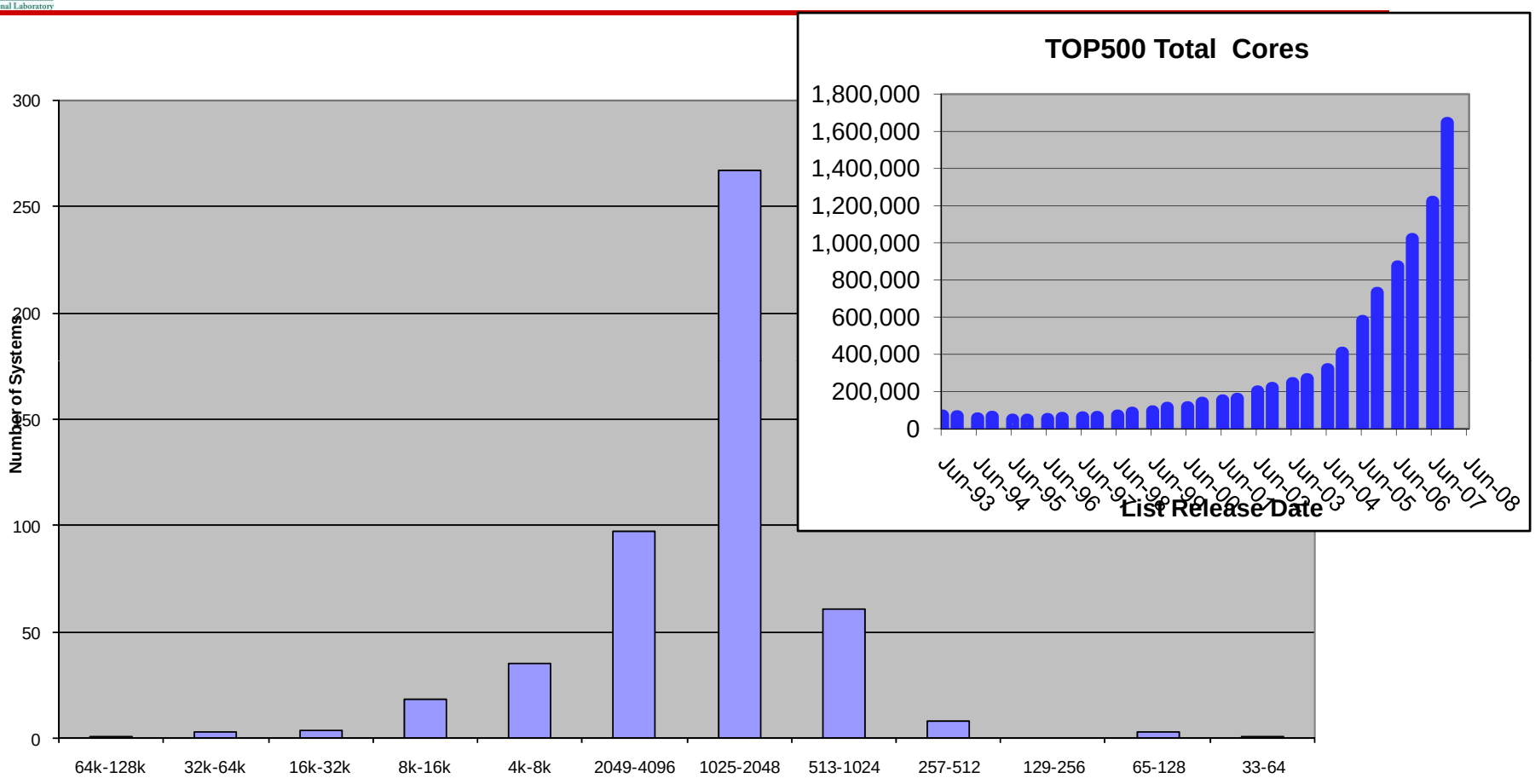
Gigabit Ethernet (270)

N/A

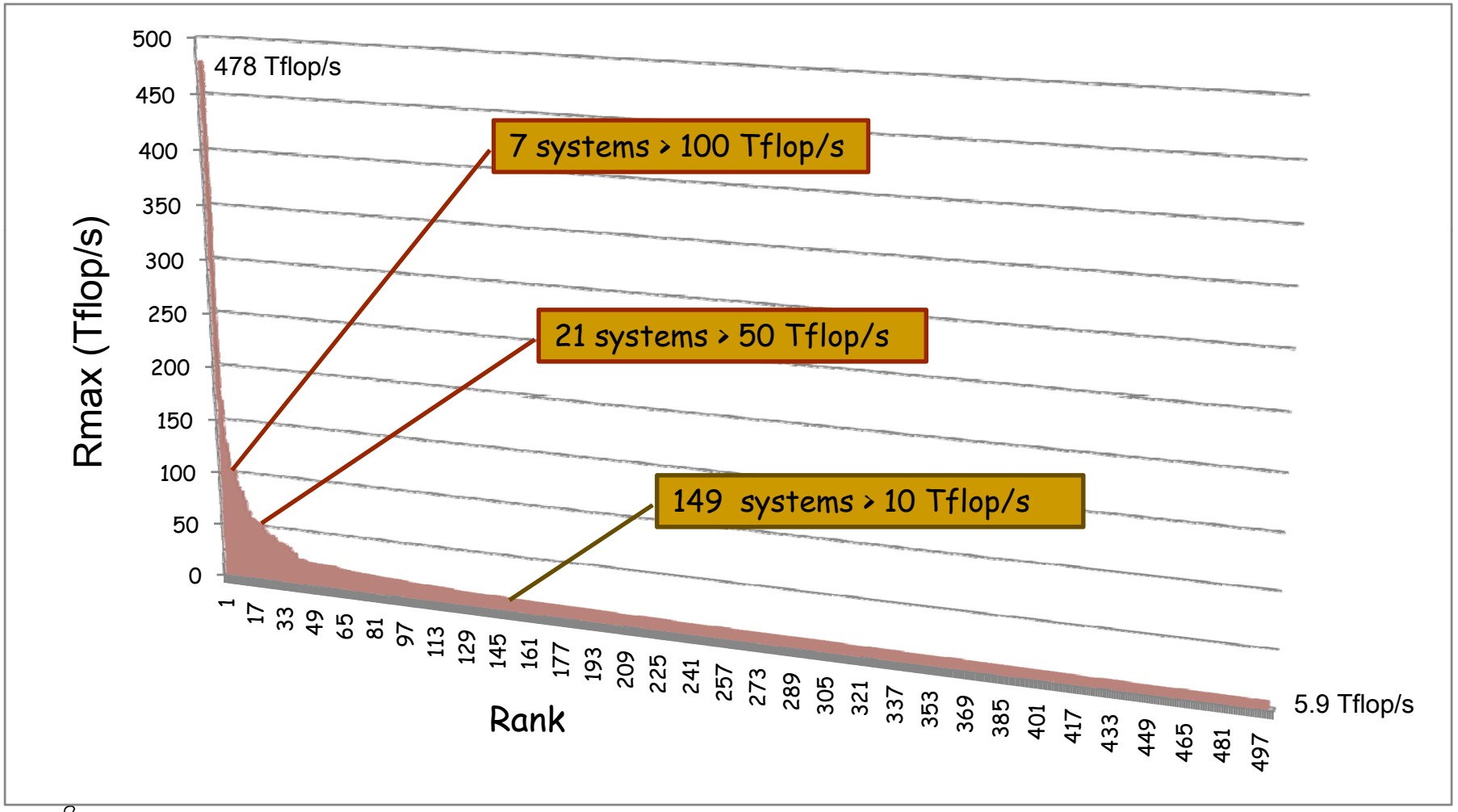
07

GigE + Infiniband + Myrinet = 82%

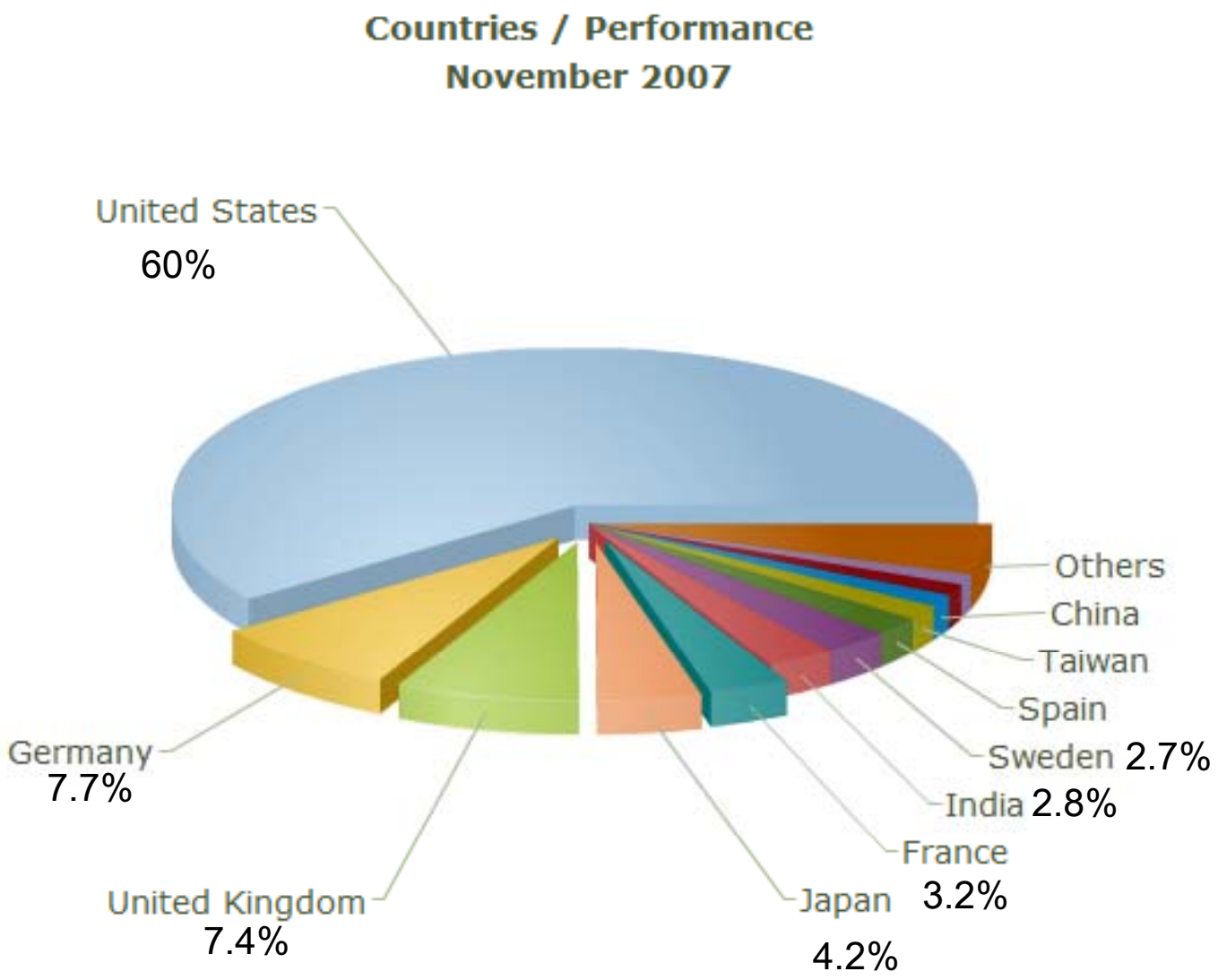
Cores per System – November 2007



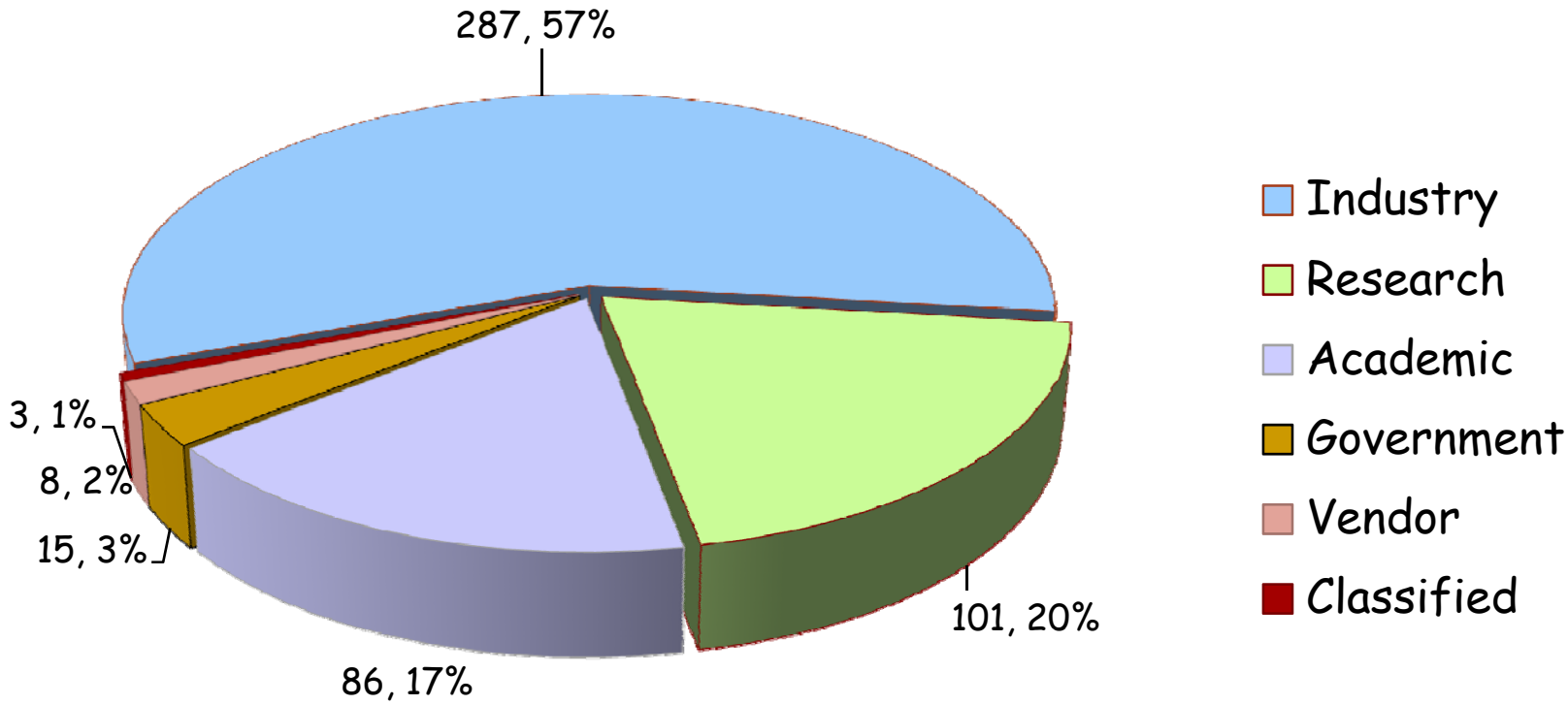
Top500 Systems November 2007



Countries / Performance (Nov 2007)



Top500 by Usage

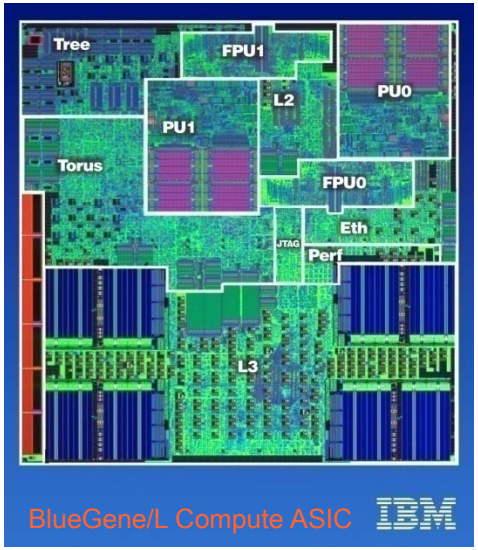


From the 30th Edition: The TOP10

	Manufacturer	Computer	Rmax [TF/s]	Installation Site	Country	Year	#Cores
1	IBM	Blue Gene/L eServer Blue Gene Dual Core 0.7 GHz	478	DOE Lawrence Livermore Nat Lab	USA	2007 Custom	212992
2	IBM	Blue Gene/P Quad Core 0.85 GHz	167	Forschungszentrum Juelich	Germany	2007 Custom	26544
3	SGI	Altix ICE 8200 Xeon Quad Core 3.0 GHz	127	SGI/New Mexico Computing Applications Center	USA	2007 Hybrid	40960
4	HP	Cluster Platform Xeon Dual Core 3.0 GHz	118	Computational Research Laboratories, TATA SONS	India	2007 Commod	12208
5	HP	Cluster Platform Dual Core 2.66 GHz	102.8	Government Agency	Sweden	2007 Commod	10240
6	Cray	Opteron Dual Core 2.4 GHz	102.2	DOE Sandia Nat Lab	USA	2007 Hybrid	9024
7	Cray	Opteron Dual Core 2.6 GHz	101.7	DOE Oak Ridge National Lab	USA	2006 Hybrid	9968
8	IBM	eServer Blue Gene/L Dual Core 0.7 GHz	91.2	IBM Thomas J. Watson Research Center	USA	2005 Custom	10160
9	Cray	Opteron Dual Core 2.6 GHz	85.4	DOE Lawrence Berkeley Nat Lab	USA	2006 Hybrid	11088
10	IBM	eServer Blue Gene/L Dual Core 0.7 GHz	82.1	Stony Brook/BNL, NY Center for Computational Sciences	USA	2006 Custom	10424

IBM BlueGene/L #1 212,992 Cores

Total of 26 systems all in the Top176



2.6 MWatts (2600 homes)
70,000 ops/s/person

(104 racks, 104x32x32)
212992 procs

Rack
(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)

180/360 TF/s
32 TB DDR

2.9/5.7 TF/s
0.5 TB DDR

90/180 GF/s
16 GB DDR

5.6/11.2 GF/s
1 GB DDR

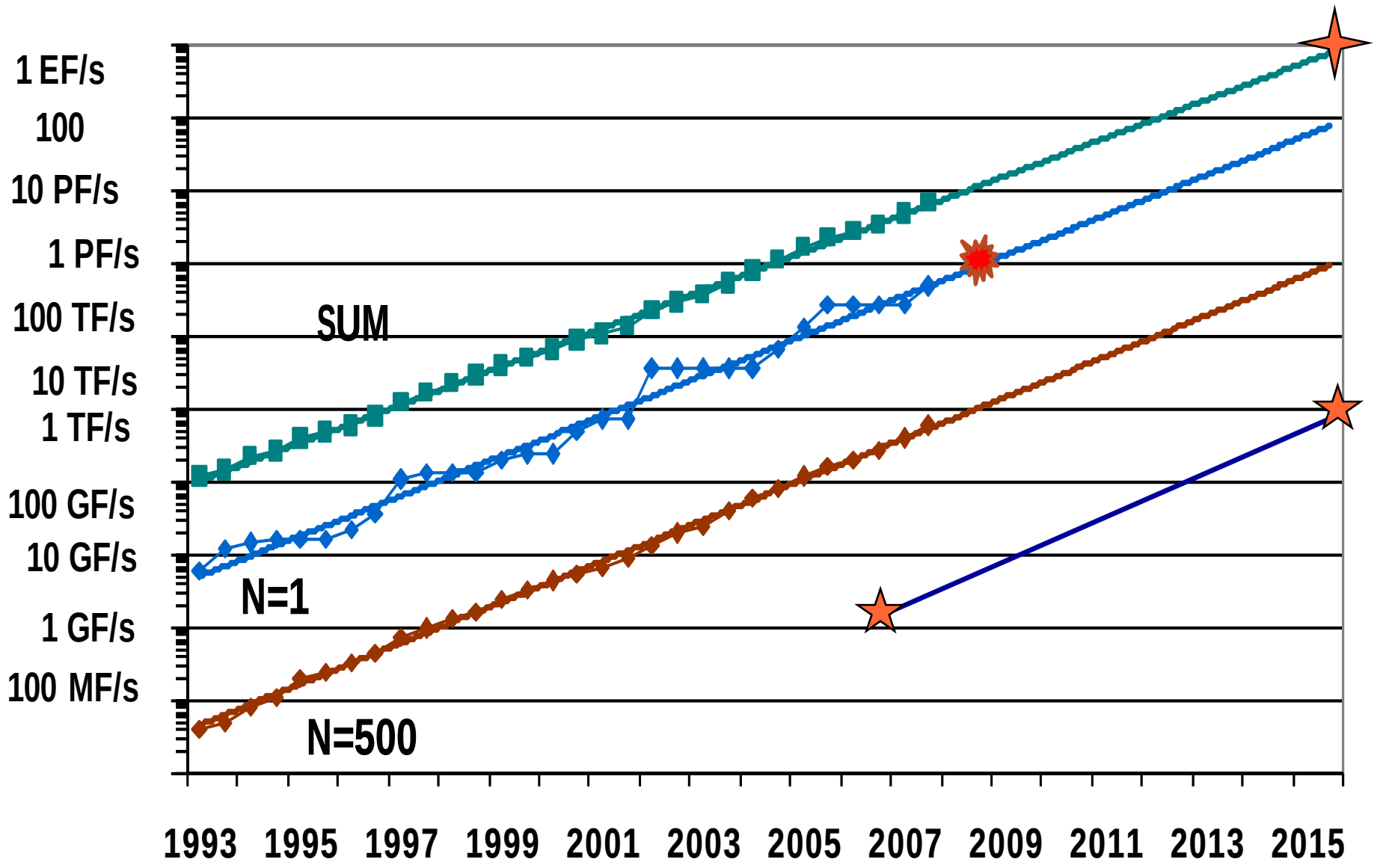
2.8/5.6 GF/s
4 MB (cache)

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). (20.7K sec about 5.7hours; n=2.5M)

"Fastest Computer"
BG/L 700 MHz 213K proc
104 racks
Peak: 596 Tflop/s
Linpack: 498 Tflop/s
84% of peak

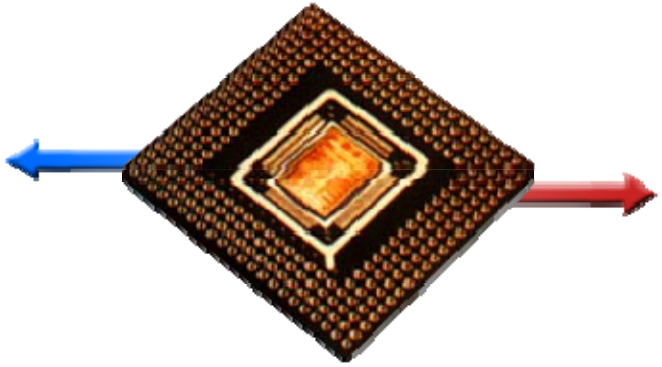
Performance Projection



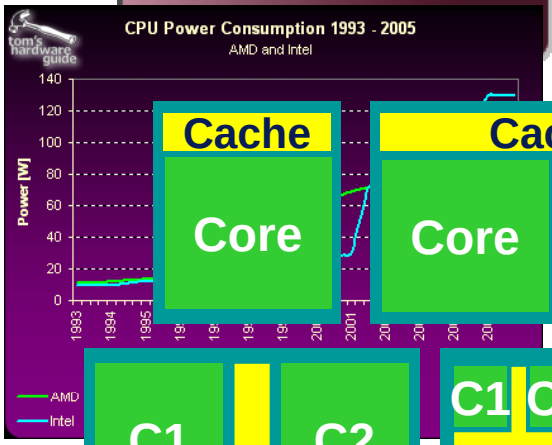
Increasing CPU Performance: A Delicate Balancing Act

Increasing the number of gates into a tight knot and decreasing the cycle time of the processor

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.

Power Cost of Frequency

- Power ∝ Voltage^2 x Frequency (V^{2}F)
- Frequency ∝ Voltage
- Power ∝ Frequency^3

	Cores	V	Freq	Perf	Power	PE (Bops/watt)
Superscalar	1	1	1	1	1	1
"New" Superscalar	1X	1.5X	1.5X	1.5X	3.3X	0.45X

Power Cost of Frequency

- Power ∝ Voltage^2 x Frequency (V^{2}F)
- Frequency ∝ Voltage
- Power ∝ Frequency^3

	Cores	V	Freq	Perf	Power	PE (Bops/watt)
Superscalar	1	1	1	1	1	1
"New" Superscalar	1X	1.5X	1.5X	1.5X	3.3X	0.45X
Multicore	2X	0.75X	0.75X	1.5X	0.8X	1.88X

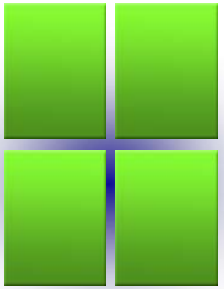
(Bigger # is better)

50% more performance with 20% less power

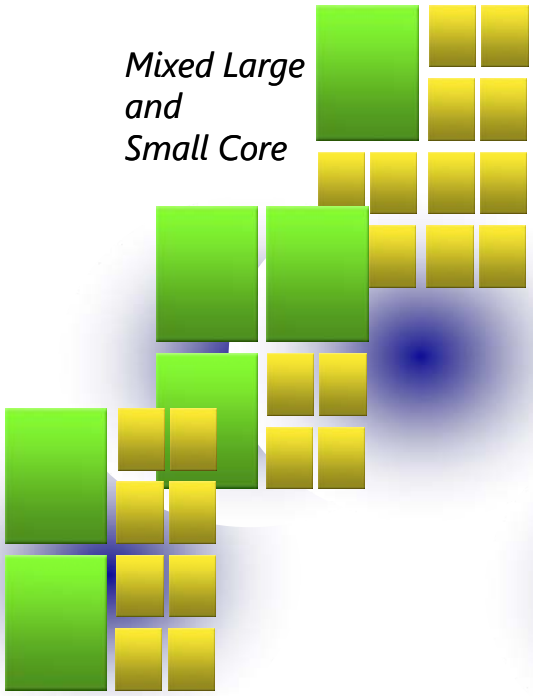
Preferable to use multiple slower devices, than one superfast device

What's Next?

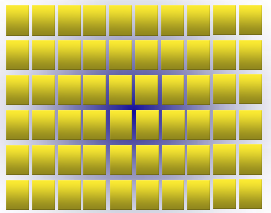
All Large Core



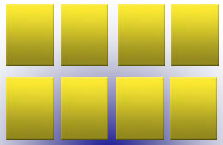
Mixed Large and Small Core



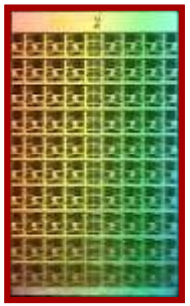
Many Small Cores



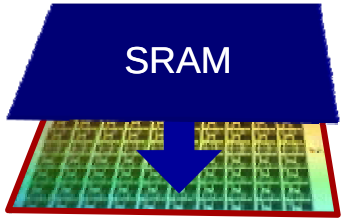
All Small Core



Many Floating-Point Cores



+ 3D Stacked Memory

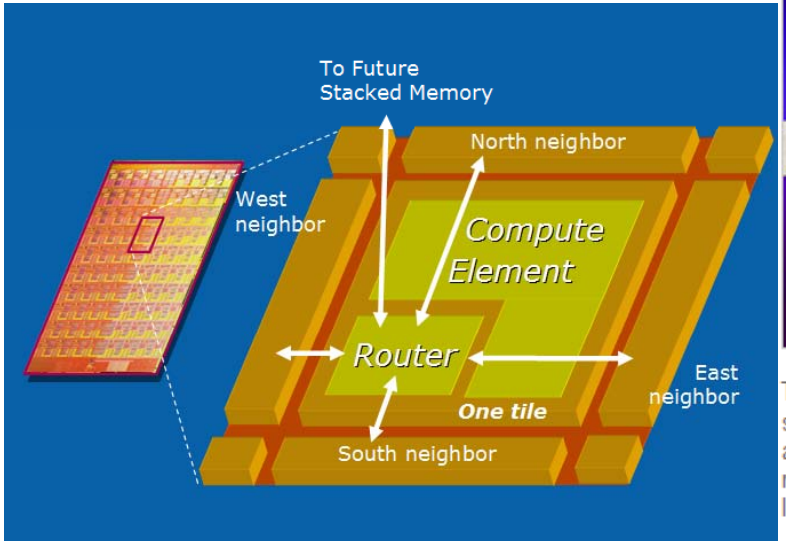


Different Classes of Chips

- Home
- Games / Graphics
- Business
- Scientific

80 Core

- **Intel's 80 Core chip**
  - 1 Tflop/s
  - 62 Watts
  - 1.2 TB/s internal BW

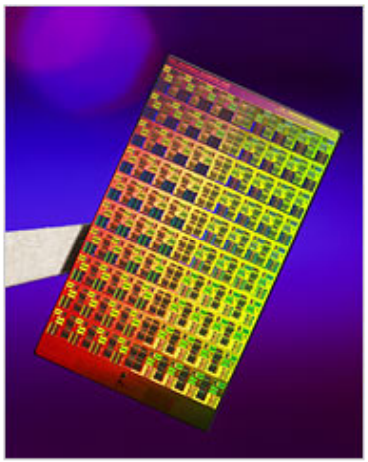


The New York Times
Technology

Intel Prototype May Herald a New Age of Processing

By JOHN MARKOFF
Published: February 12, 2007

SAN FRANCISCO, Feb. 11 — Intel will demonstrate on Monday an experimental computer chip with 80 separate processing engines, or cores, that company executives say provides a model for commercial chips that will be used widely in standard desktop, laptop and server computers within five years.



Intel

The Teraflop Chip has 80 separate processing engines and takes advantage of manufacturing technology that Intel introduced last month.

The new processor, which the company first described as a Teraflop Chip at a conference last year, will be detailed in a technical paper to be presented on the opening day of the International Solid States Circuits Conference, beginning here on Monday.

While the chip is not compatible with Intel's current chips, the company said it had already begun design work on a commercial version that would essentially have dozens or even hundreds of Intel-compatible microprocessors laid out in a tiled pattern on a single chip.

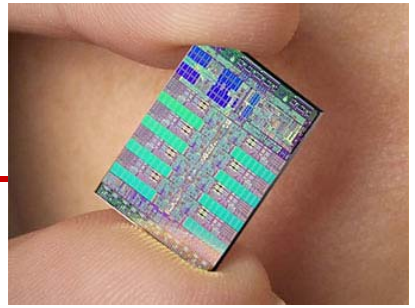
Constantly Evolving - Hybrid Design

- More and more High Performance Computers will be built on a Hybrid Desing

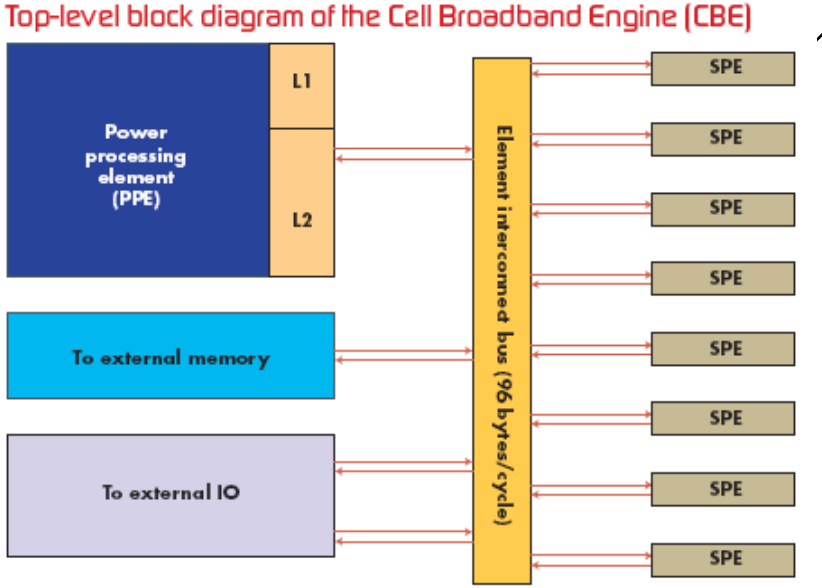
- Cluster of Cluster systems
  - Multicore nodes in a cluster
- Nodes augmented with accelerators
  - ClearSpeed, GPUs, Cell

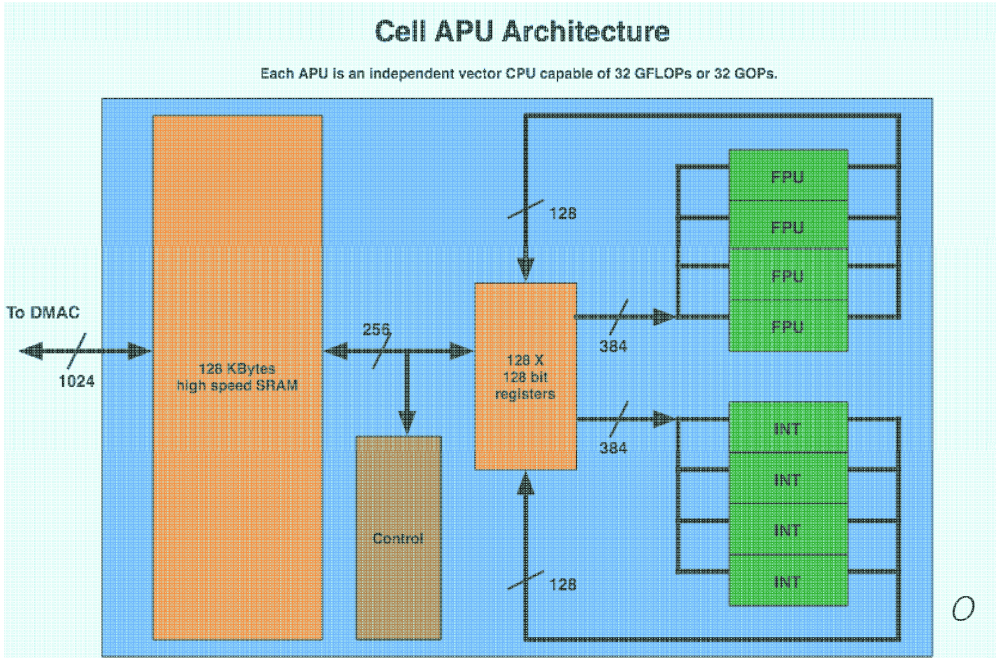
- Japanese 10 PFlop/s "Life Simulator"
  - Vector+Scalar+Grape:
    - Theoretical peak performance: >1-2 PetaFlops from Vector + Scalar System, ~10 PetaFlops from MD-GRAPE-like System
- LANL's Roadrunner
  - Multicore + specialized accelerator processors

IBM Cell Processor



- The PlayStation 3's CPU based on a "Cell" processor
- Each Cell contains 8 APUs.
  - An SPE is a self contained vector processor which acts independently from the others.
  - 4 floating point units capable of a total of 25 Gflop/s (5 Gflop/s each @ 3.2 GHz)
  - 204 Gflop/s peak! 32 bit floating point;
  - 102 Gflop/s for 64 bit floating point. (new chips, old ones at 14 Gflop/s!)





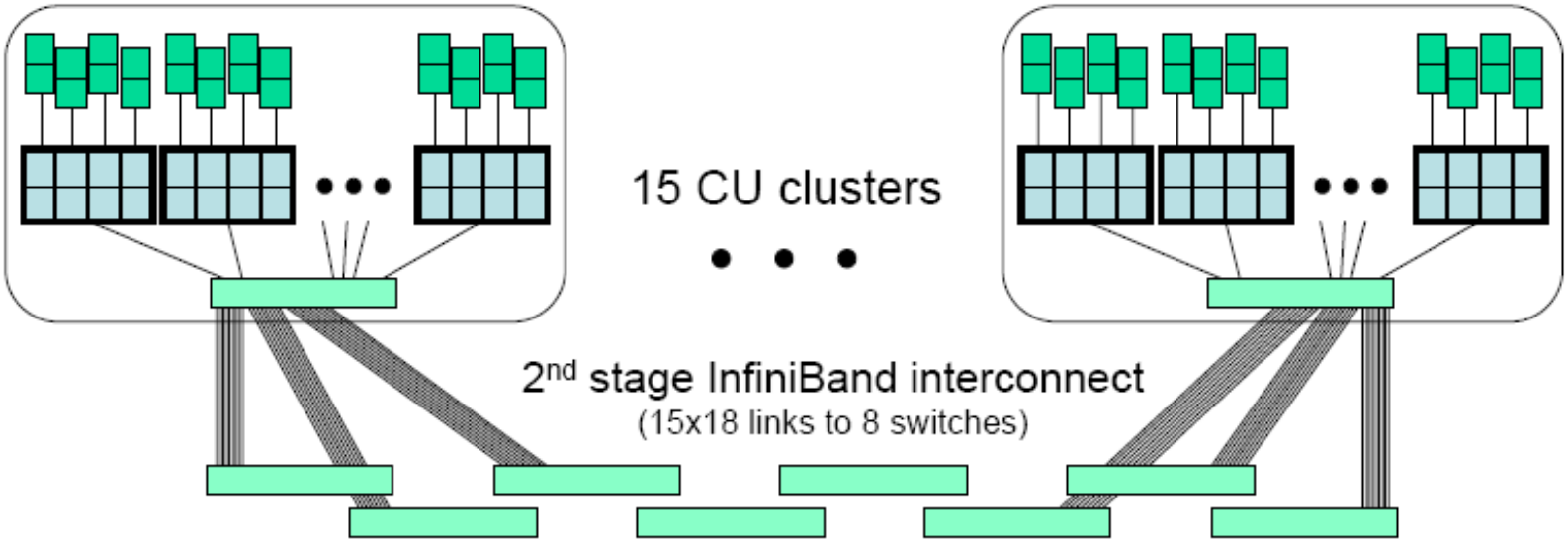
Accelerated Roadrunner

Hybrid Design

"Connected Unit" cluster
144 quad-socket
dual-core nodes
(138 w/ 4 dual-Cell blades)
InfiniBand interconnects

In aggregate:
8,640 dual-core Opterons + 16,560 eDP Cell chips
76 TeraFlops Opteron + ~1.7 PetaFlops Cell

75,600 cores



Los Alamos NATIONAL LABORATORY EST. 1943

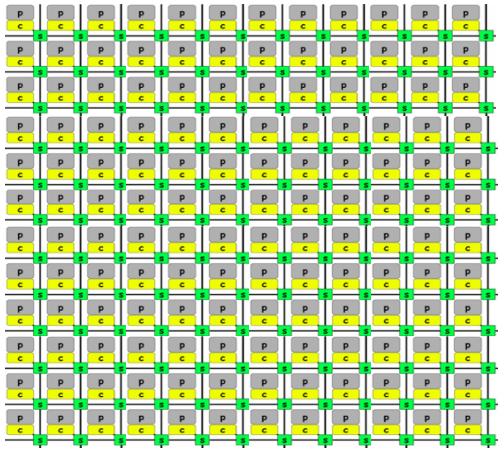
RR-14

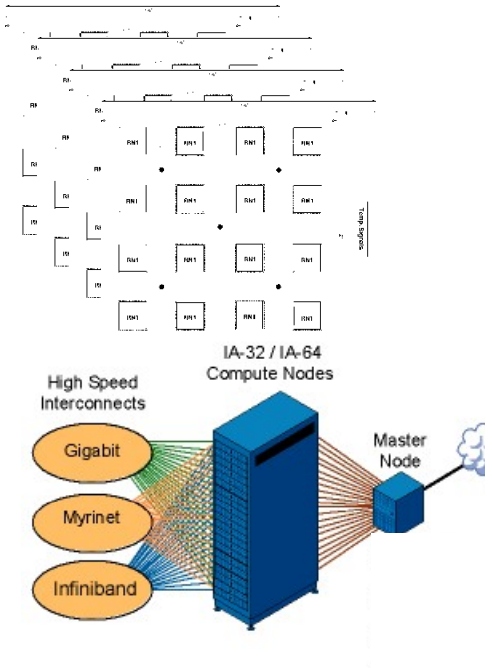
IBM

1152 AMD cores / cluster each core with a Cell processor

Future Large Systems, Say in 5 Years

- 128 cores per socket
- 32 sockets per node
- 128 nodes per system
- System = 128*32*128
  = 524,288 Cores!
- And by the way, its 4 threads of exec per core
- That's about 2M threads to manage





Collaborators

- **Top500 Team**
  - Erich Strohmaier, NERSC
  - Hans Meuer, Mannheim
  - Horst Simon, NERSC

http://www.top500