ALGORÍTMOS PARALELOS

Neyval C. Reis Jr.

OUTUBRO/2004

Laboratório de Computação de Alto Desempenho

DI/UFES

UFES
UNIVERSIDADE FEDERAL DO ESPÍRITO SANTO

Programa do Curso

1. Introdução
2. Arquitetura de Computadores
3. Arquiteturas de Sistemas Paralelos
4. Computação de Alto Desempenho
5. Programação Paralela (modelos e paradigmas)
6. Análise de Desempenho e Instrumentação
7. Aplicações

UFES
UNIVERSIDADE FEDERAL DO ESPÍRITO SANTO

Tecnologias

1. MPI – Message Passing Interface
2. HPF – High Performance Fortran
3. HLRC e ThreadMarks
4. Threads
5. OPEN-MP
6. Bibliotecas Numéricas

UFES
UNIVERSIDADE FEDERAL DO ESPÍRITO SANTO

Por que

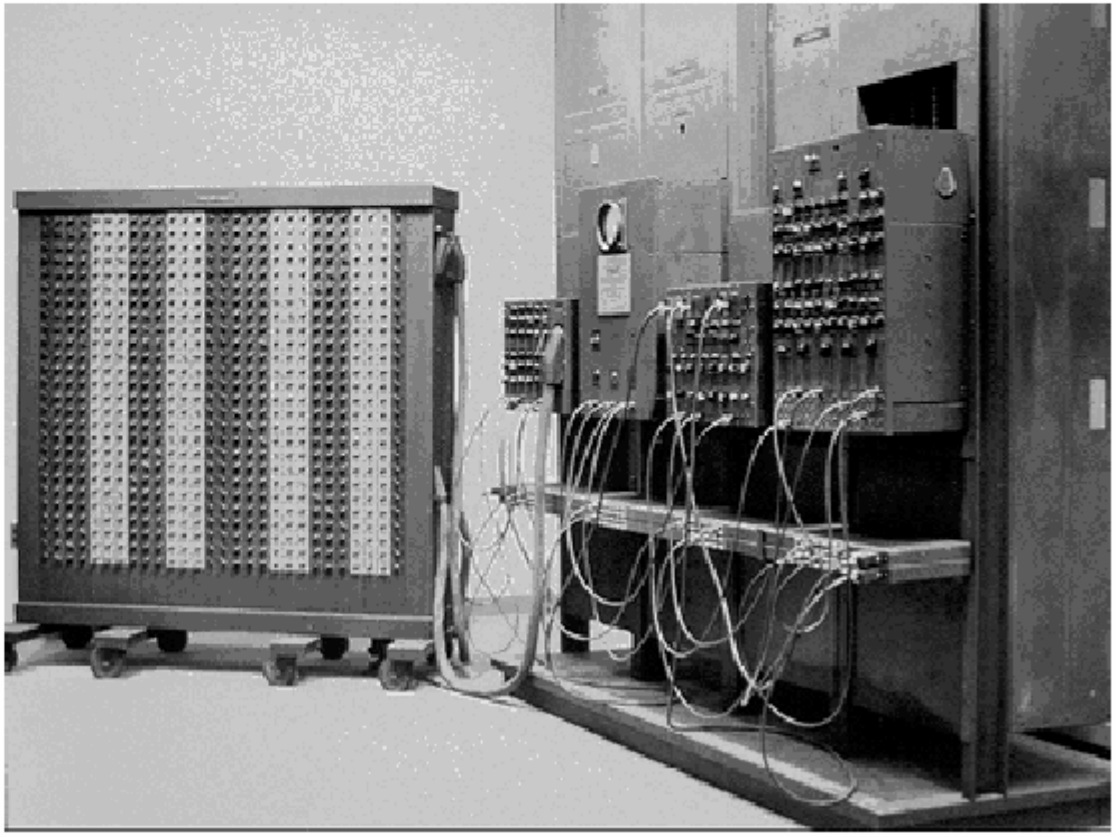
Processamento

Paralelo ?

Beginnings of Computing

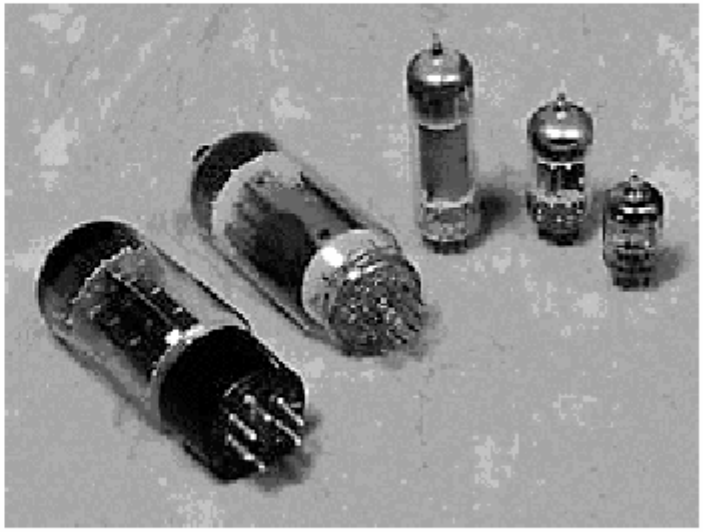
WW II - ENIAC (Electronic Numerical Integrator And Computer)

- Developed for calculations of ballistic trajectories
- Programs were entered by hardwiring



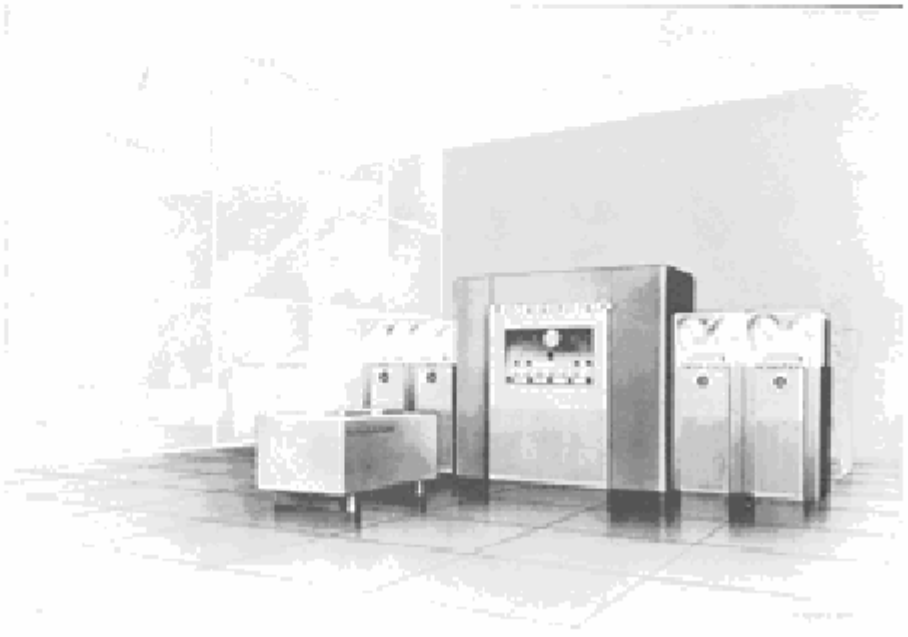
Early History of Computing: 1950s

- **During this time:**
  - computers were expensive and difficult to use
  - access was restricted (military use)
- **MANIAC 1952**
  - First program stored internally
  - Used to design H bomb
  - 2300 vacuum tubes



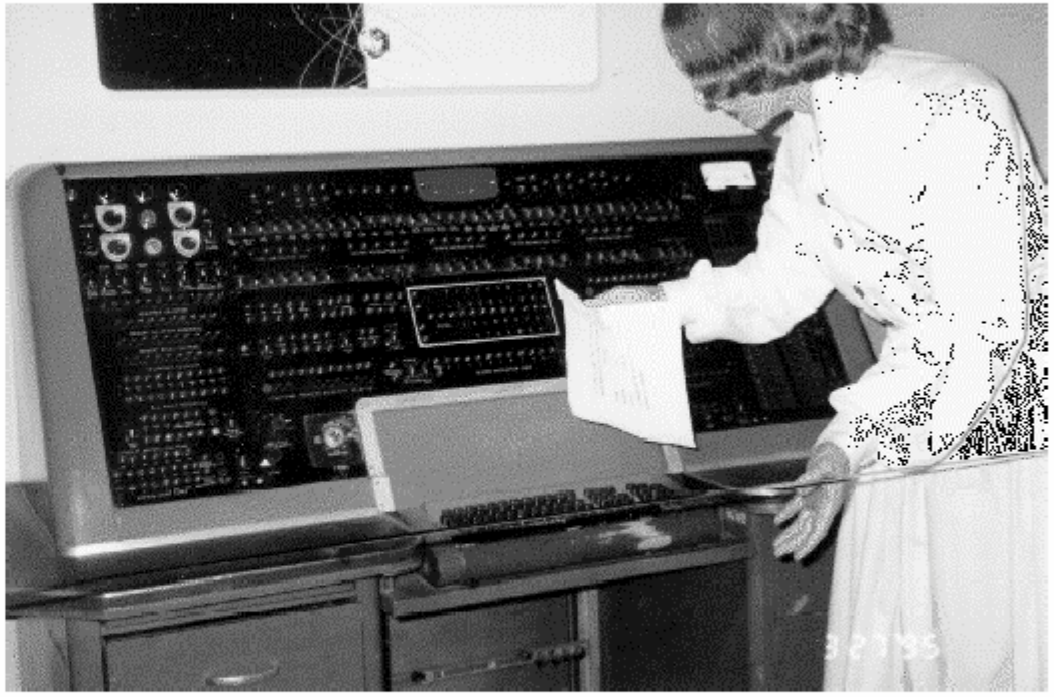
Early History of Computing: 1950s

- **UNIVAC 1951**
  - Developed by Army, Census Bureau, NBS
  - 43 total machines were produced



The UNIVAC

- 5000 vacuum tubes
- First magnetic tape storage



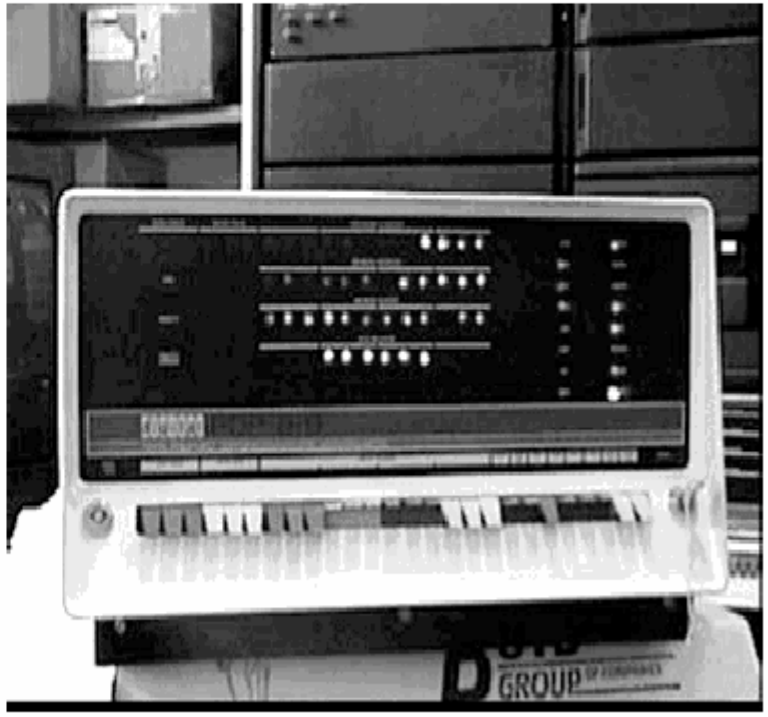
By the Mid to late 50s:

– Fewer than 1000 computers exist in US.

Computing in the 60s

Important developments:

- Transistors replaced vacuum tubes.
- First mass-produced minicomputer (DEC PDP-8) released in 1965.
- UNIX operating system developed in 1969.

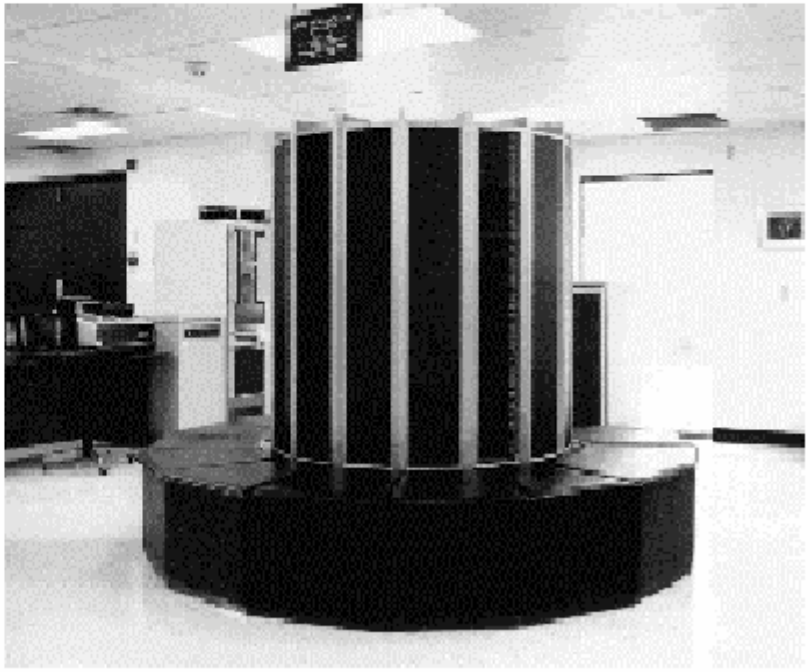


DEC PDP-8

Computing in the 70s

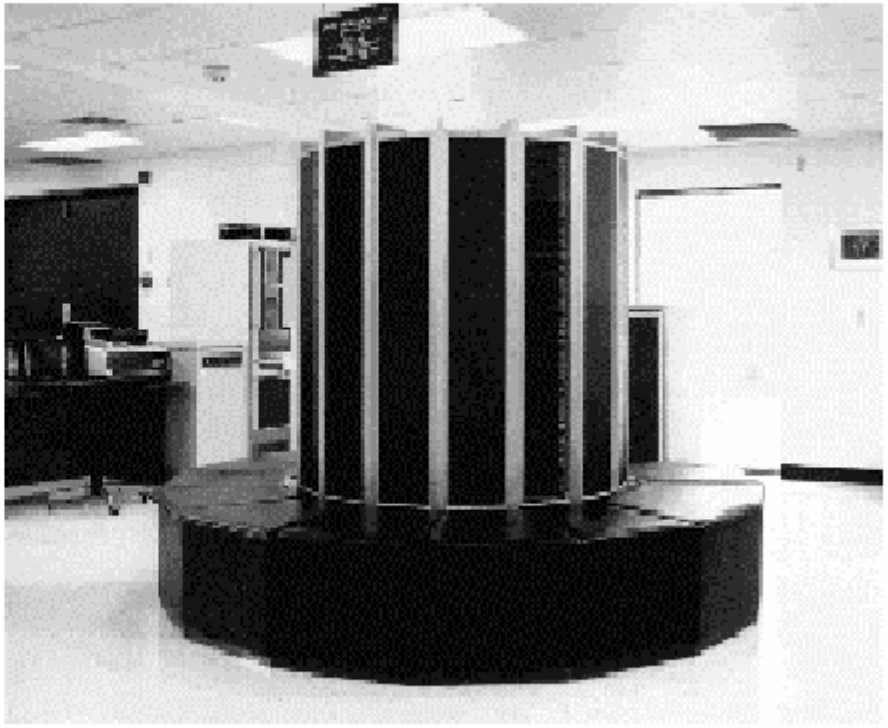
Important developments:

- Microprocessor chip introduced by Intel in 1972.
- First supercomputer developed in 1976.



The Development of the Supercomputer

- CRAY-1 was developed in 1976.
- Cray-1 had a speed of 100 Mflops.
- Today's Intel Pentium 4 (2.5 GHz) has speed of 5000 Mflops.

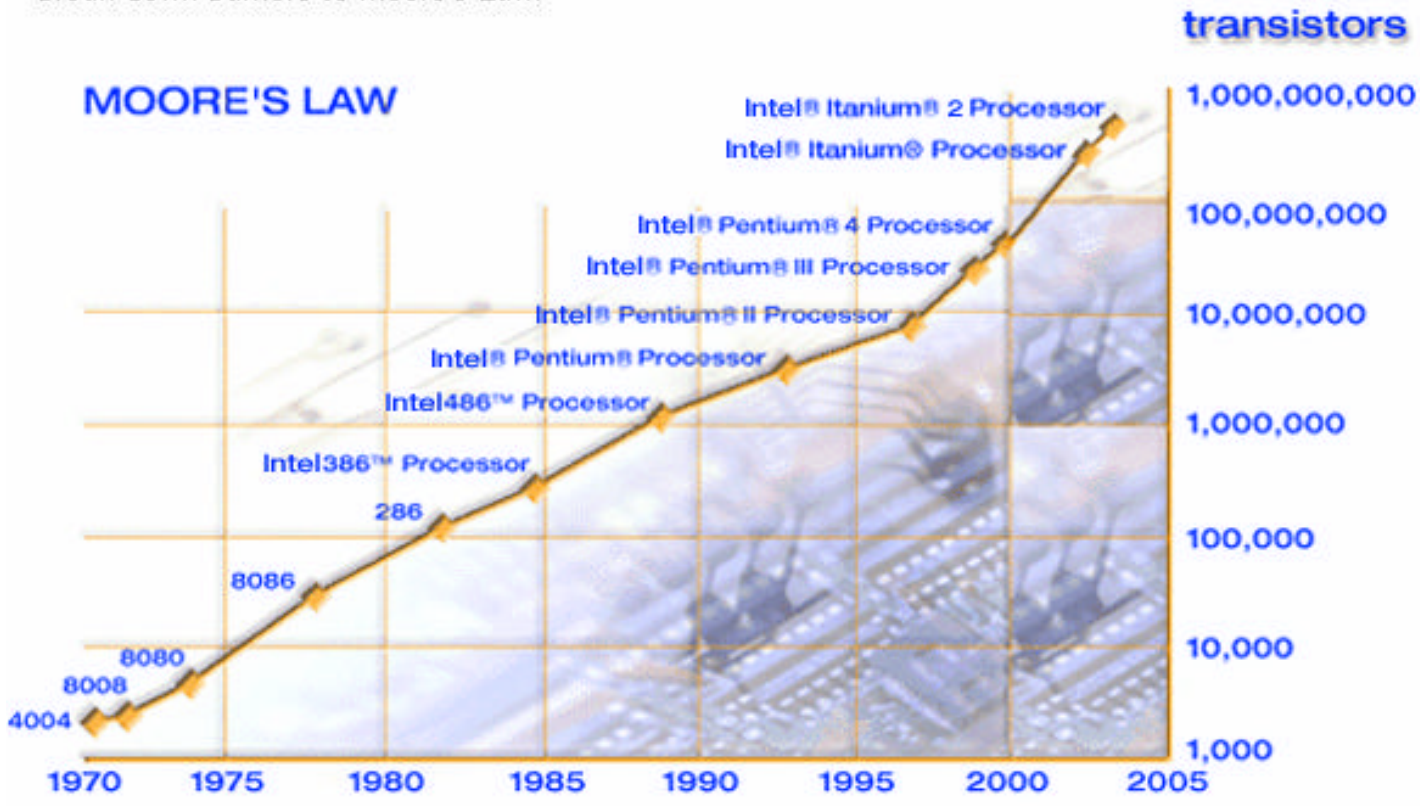


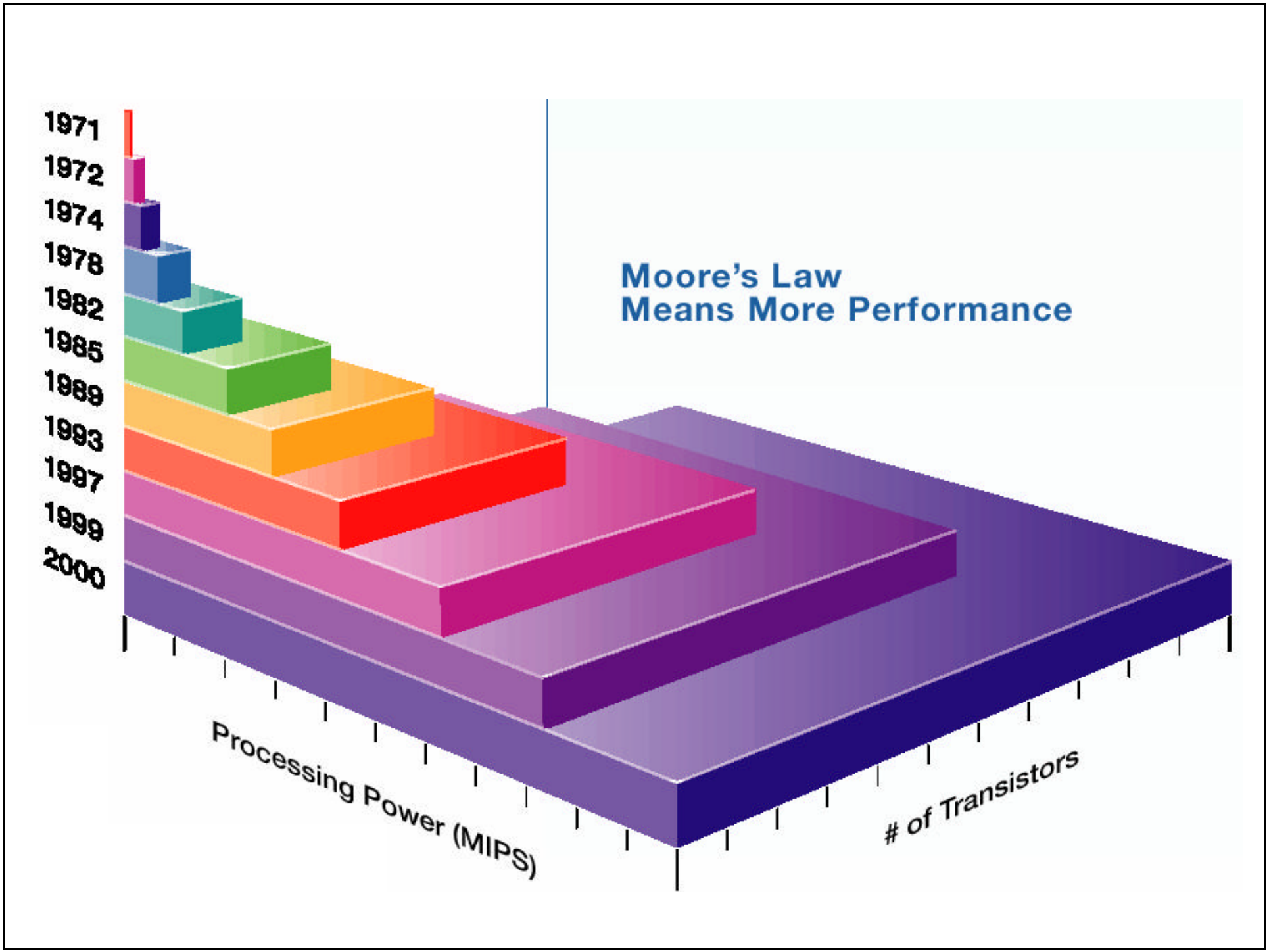
Gordon Moore

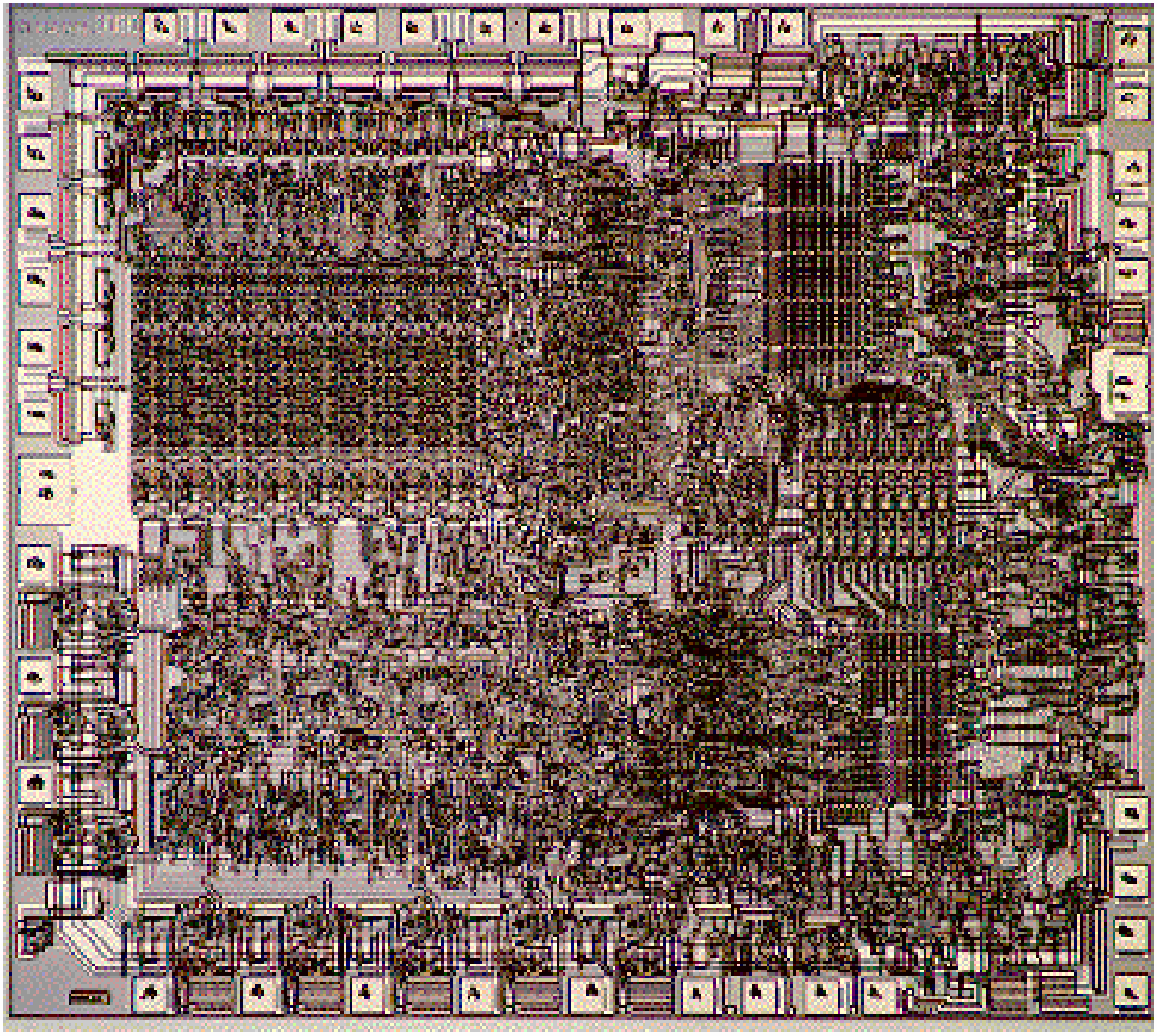
Gordon Moore made his famous observation in 1965, just four years after the first planar integrated circuit was discovered. The press called it "Moore's Law" and the name has stuck. In his original paper, Moore observed an exponential growth in the number of transistors per integrated circuit and predicted that this trend would continue. Through Intel's relentless technology advances, Moore's Law, the doubling of transistors every couple of years, has been maintained, and still holds true today. Intel expects that it will continue at least through the end of this decade. The mission of Intel's technology development team is to continue to break down barriers to Moore's Law.

Focus Research

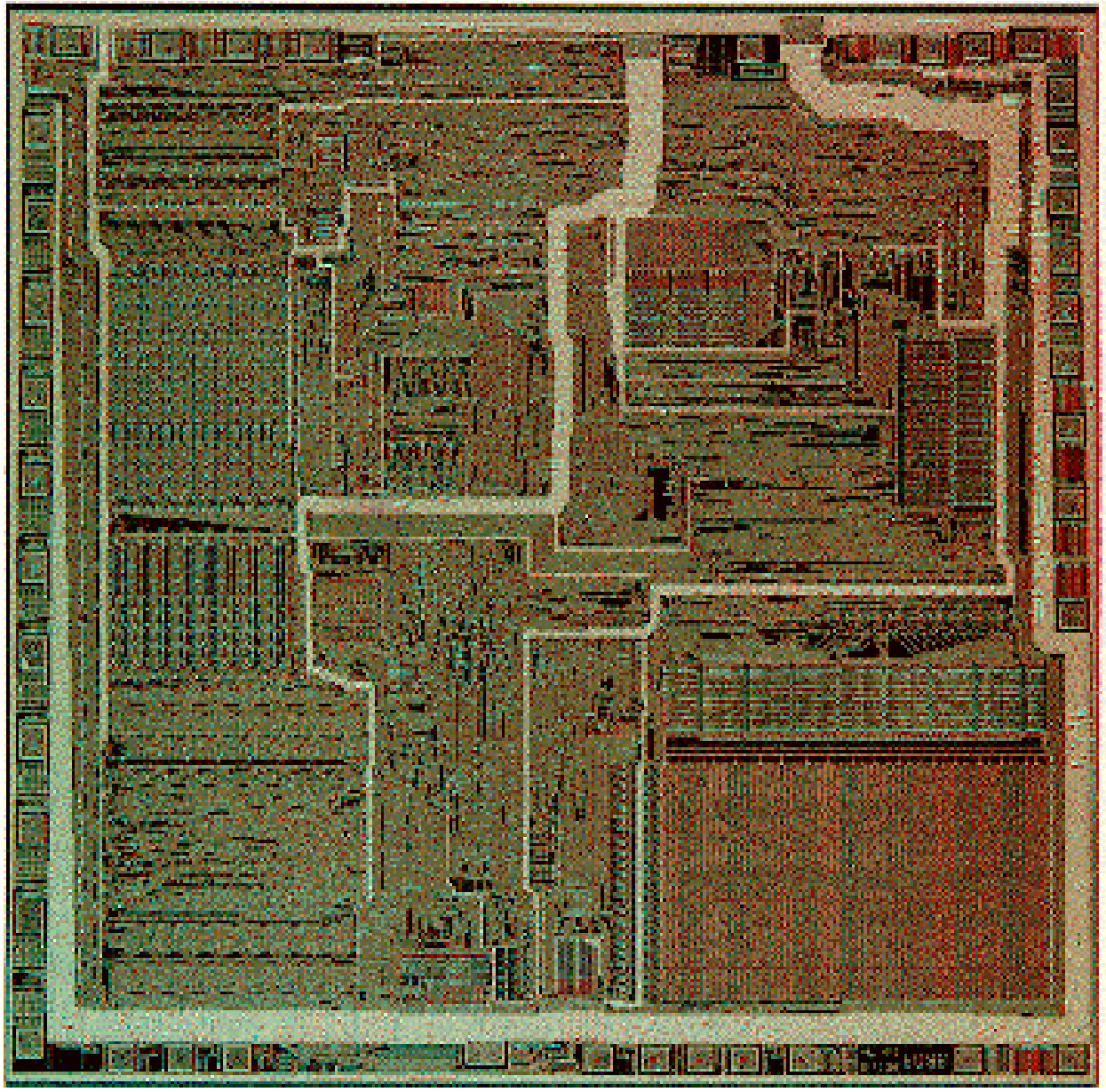
- Logic Processes
- Nanotechnology
- Lithography
- 300 mm Wafers
- Flash & Other Non-Volatile Memory Technologies
- Packaging



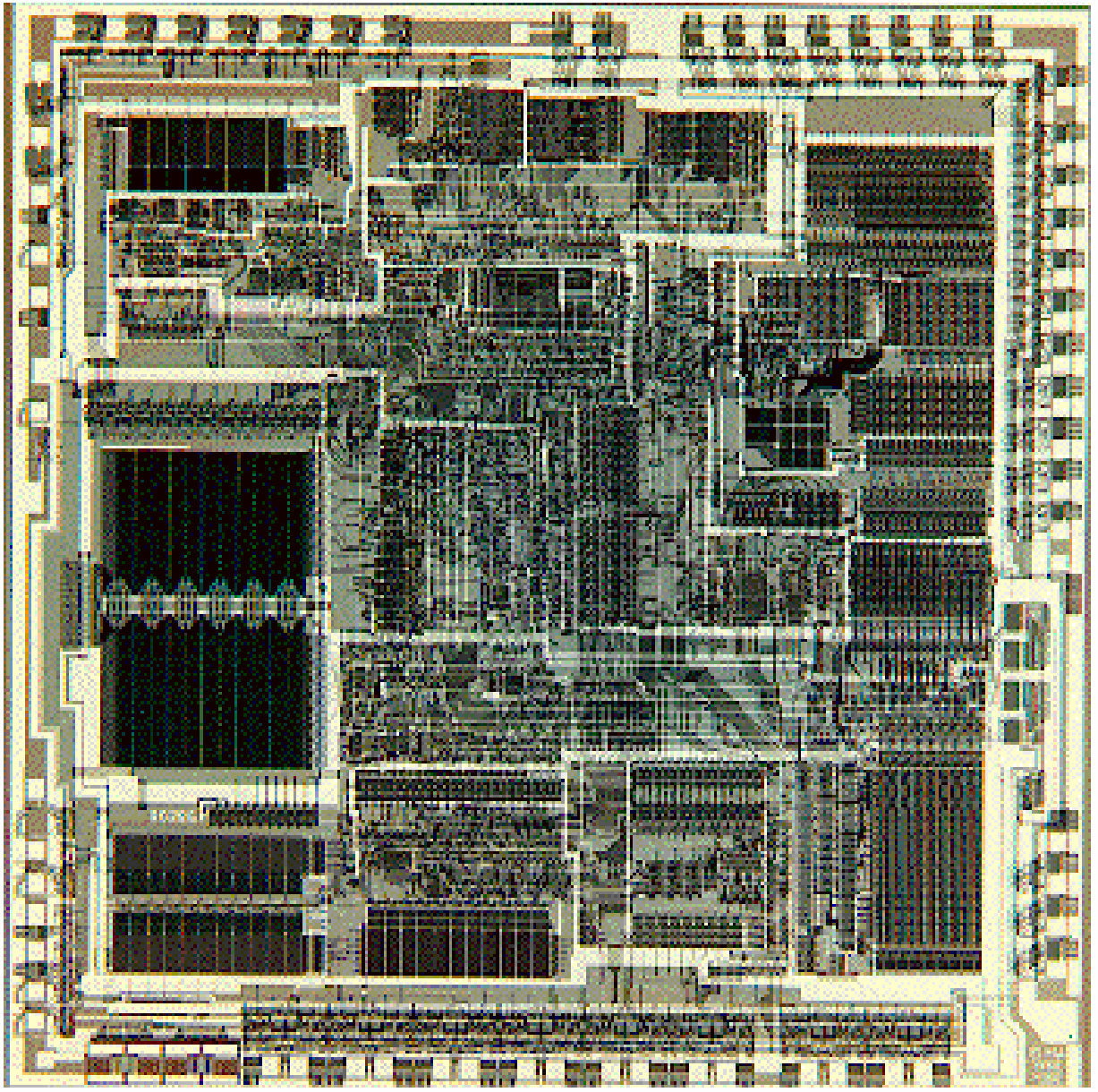

1971
1972
1974
1978
1982
1985
1989
1993
1997
1999
2000
Moore's Law
Means More Performance
Processing Power (MIPS)
# of Transistors



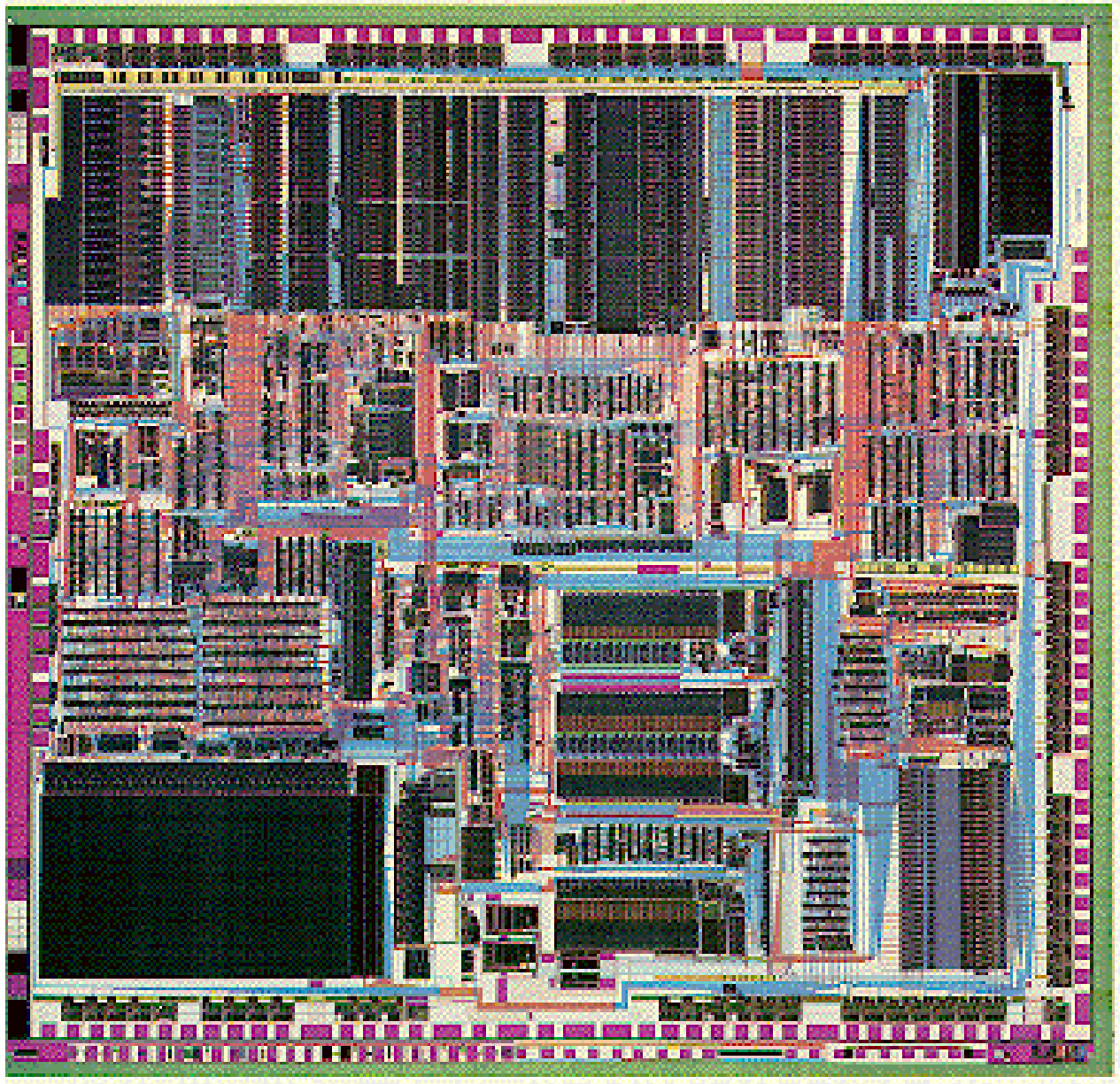
Intel 8080, 1975, 4,500 transistors



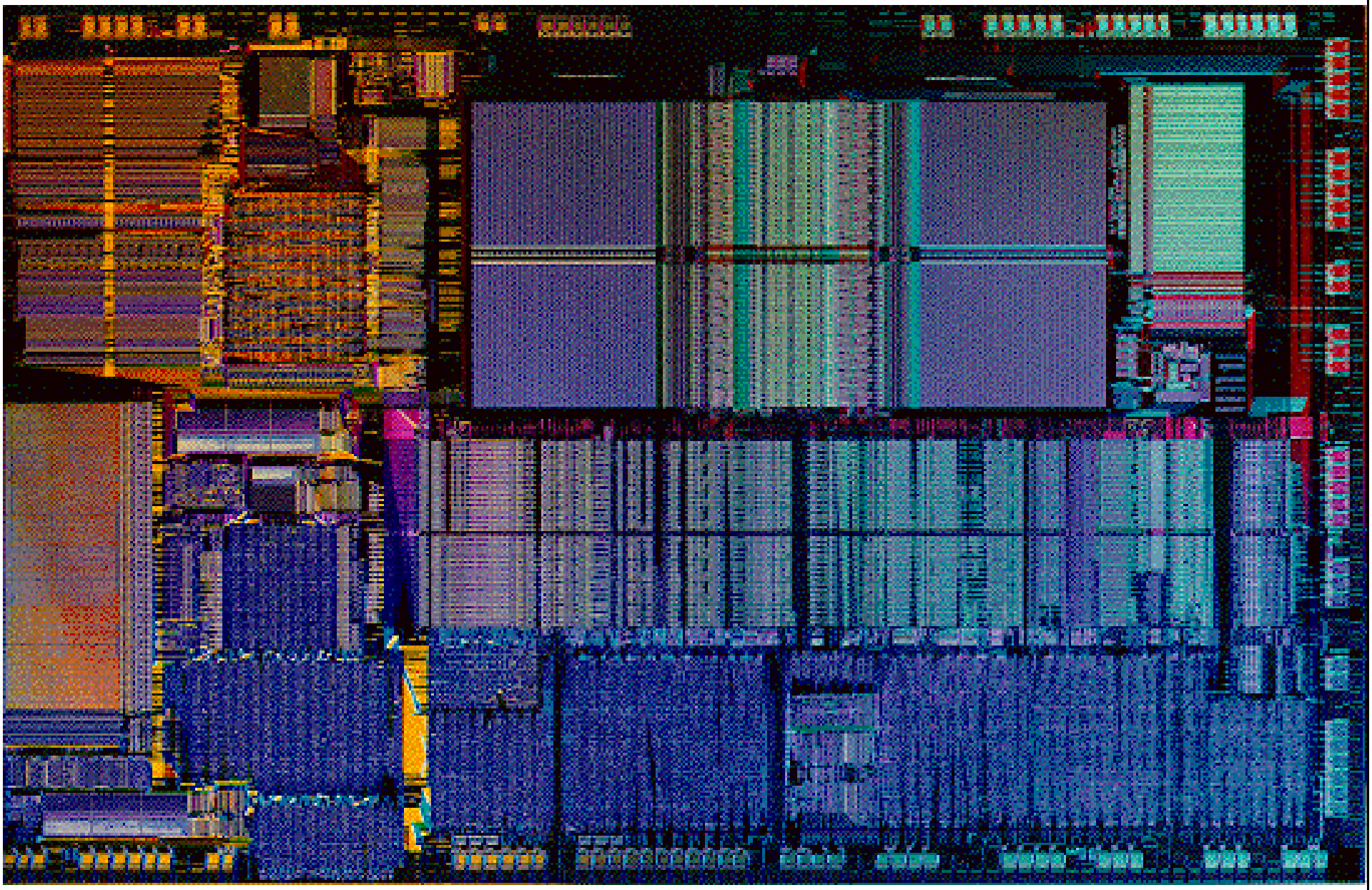
Intel 8086, 1978, 29,000 transistors



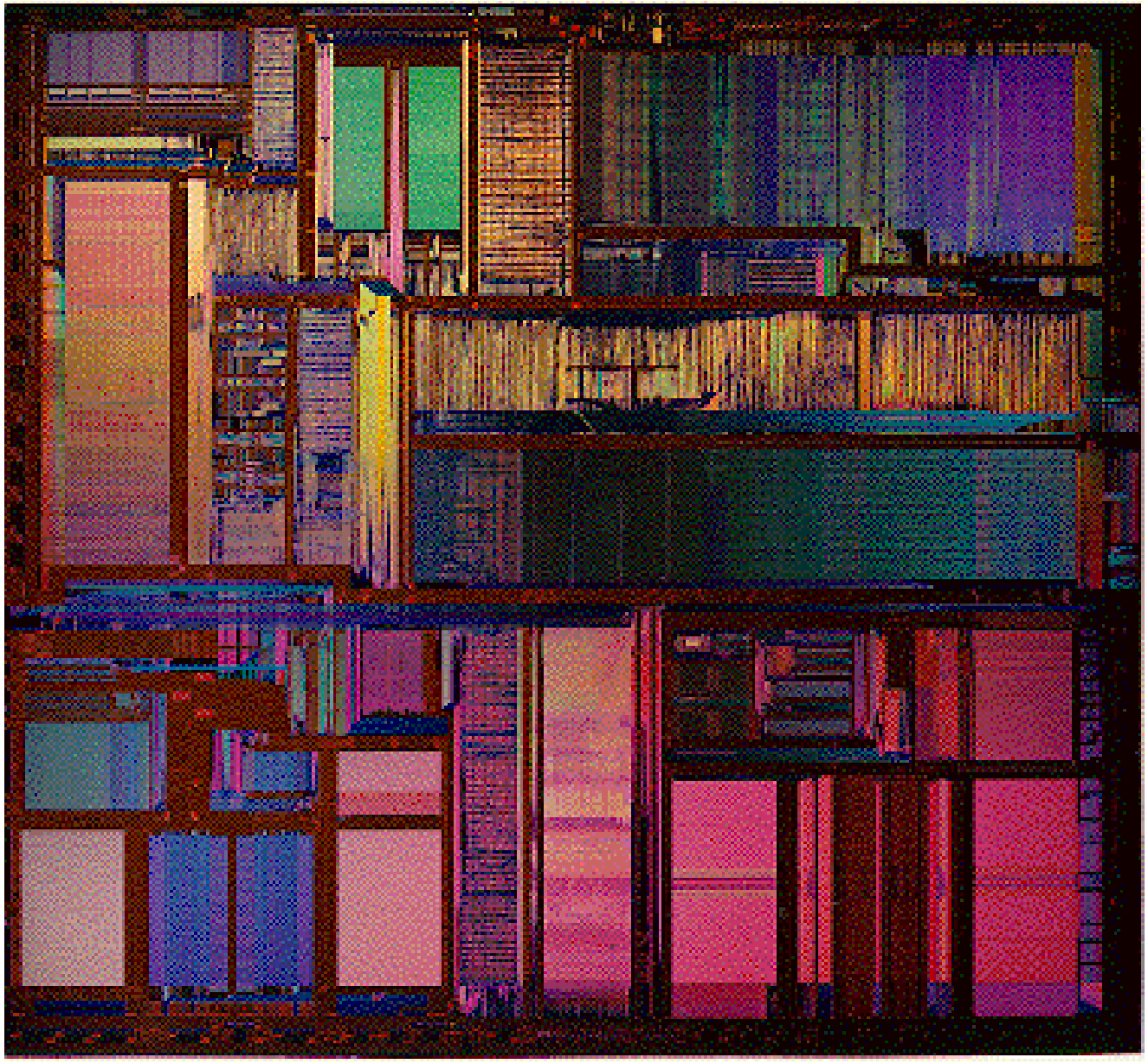
Intel 80286, 1982, 90,000 transistors



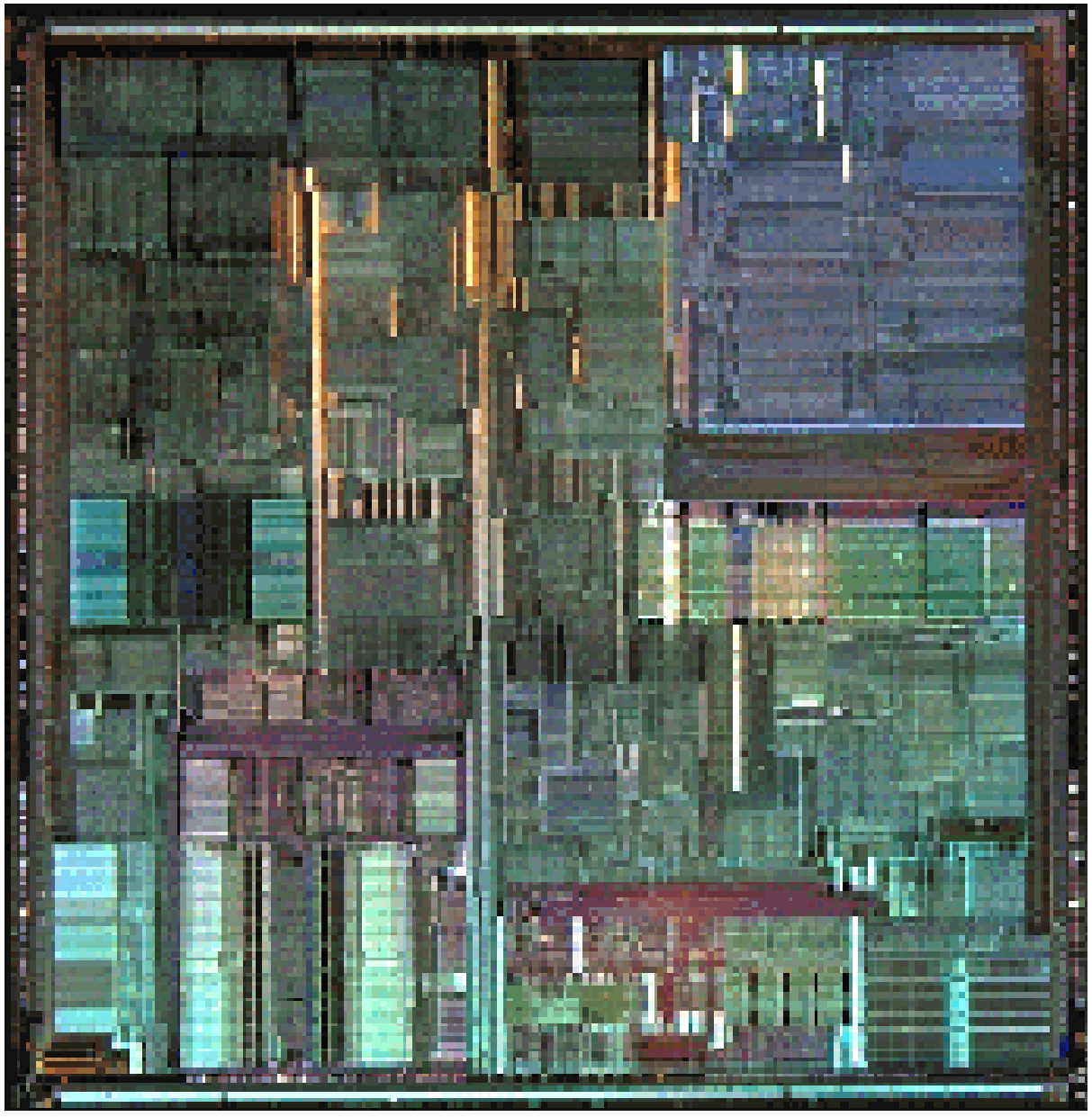
Intel 386, 1985, 229,000 transistors



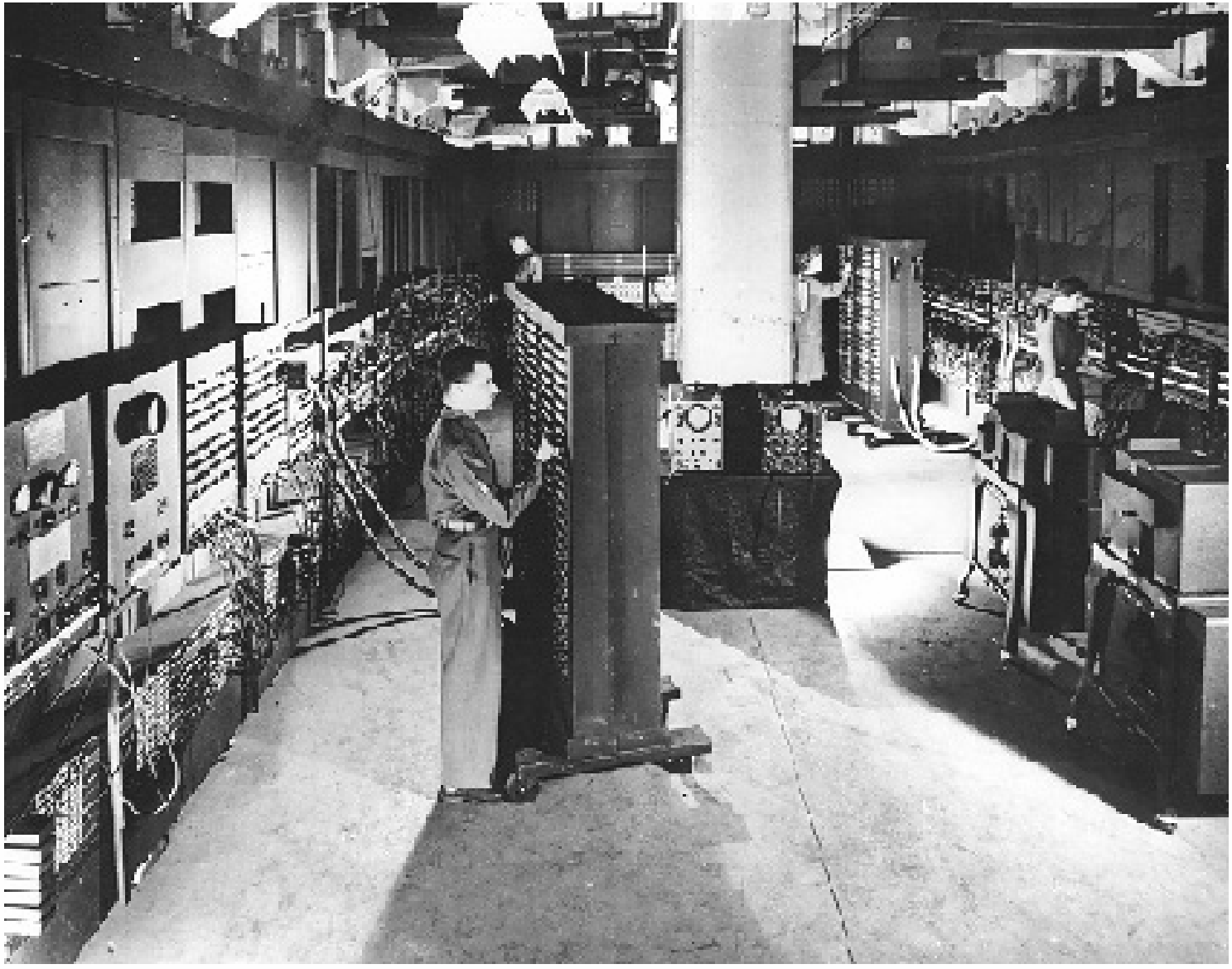
Intel 486, 1989, 1.2 million transistors



Intel Pentium, 1993, 3.1 million transistors



Intel Pentium Pro, 1995, 5.5 million transistors



ENIAC: February 14 1946, 17,468 vacuum tubes, 5,000 additions/second

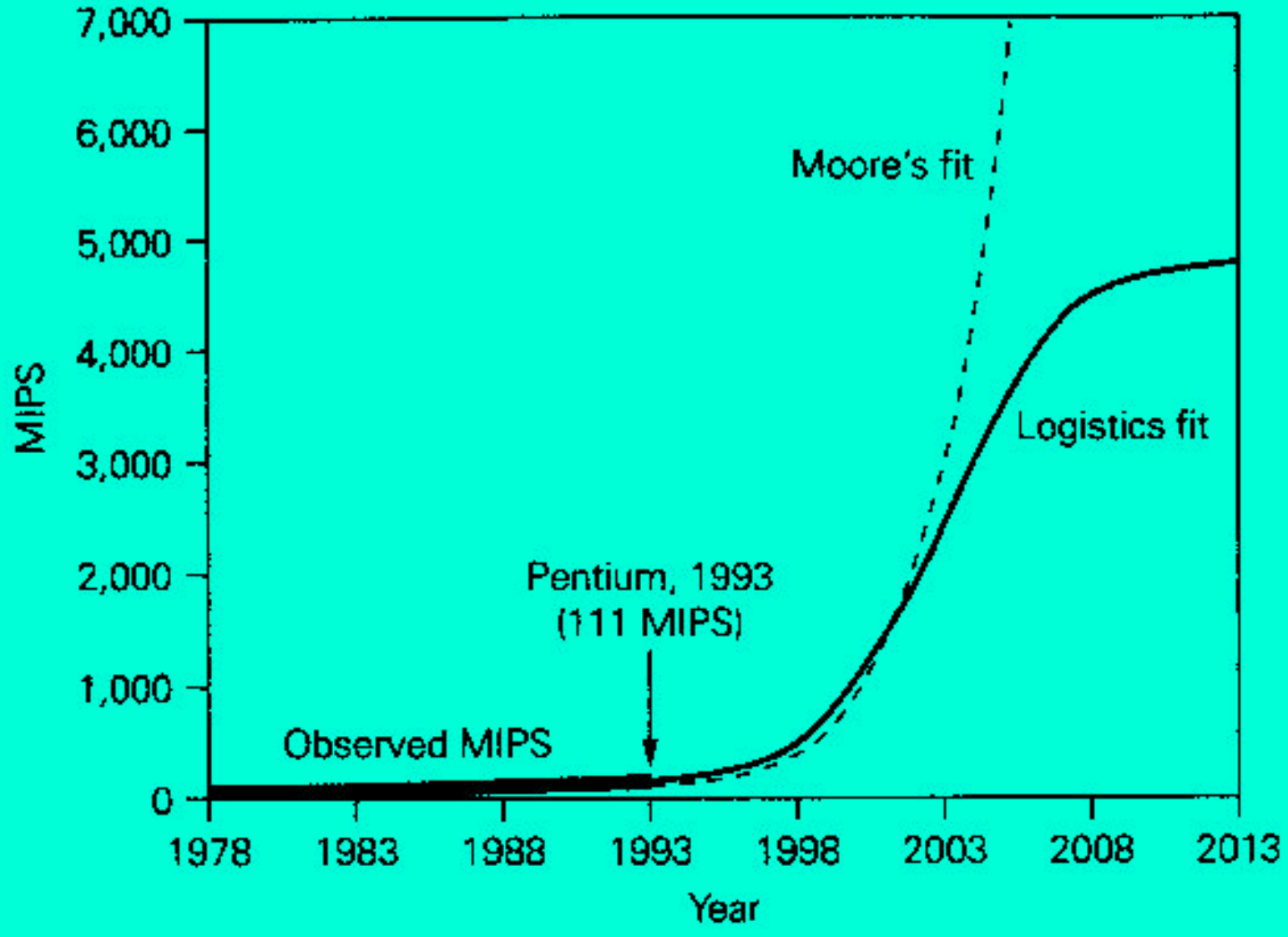
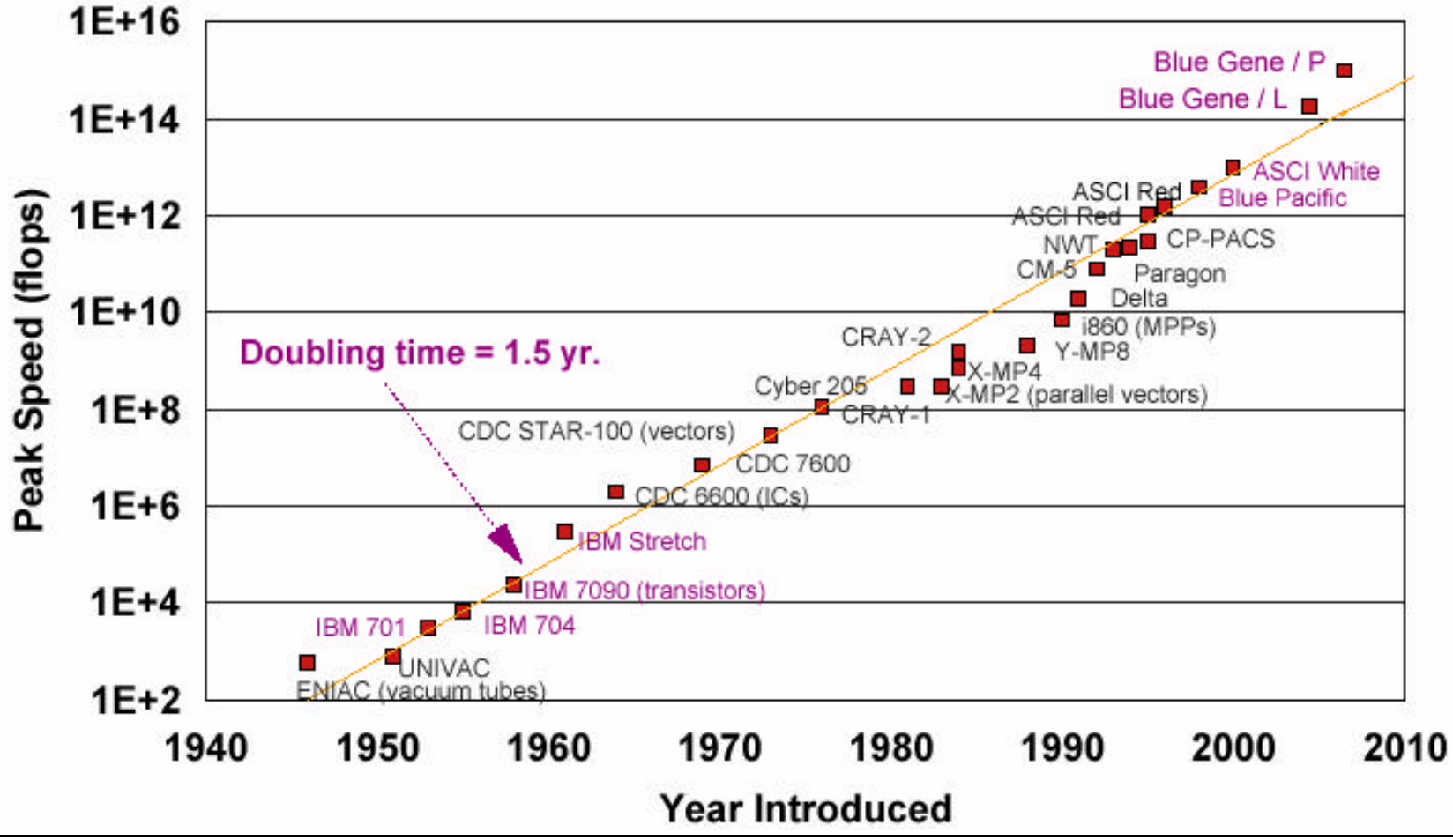


Figure 2.1 Single-processor performance in MIPS as observed, predicted by Moore's Law, and forecast by the authors.

Supercomputer Peak Performance



Perspectivas

para o Futuro

www.top500.org

UFES
UNIVERSIDADE FEDERAL DO ESPIRITO SANTO

PRESENTED BY
UNIV. OF MANNHEIM
UNIV. OF TENNESSEE
NERSC/LBNL

SUBMIT YOUR SITE

SEARCH FOR:

List for June 2004

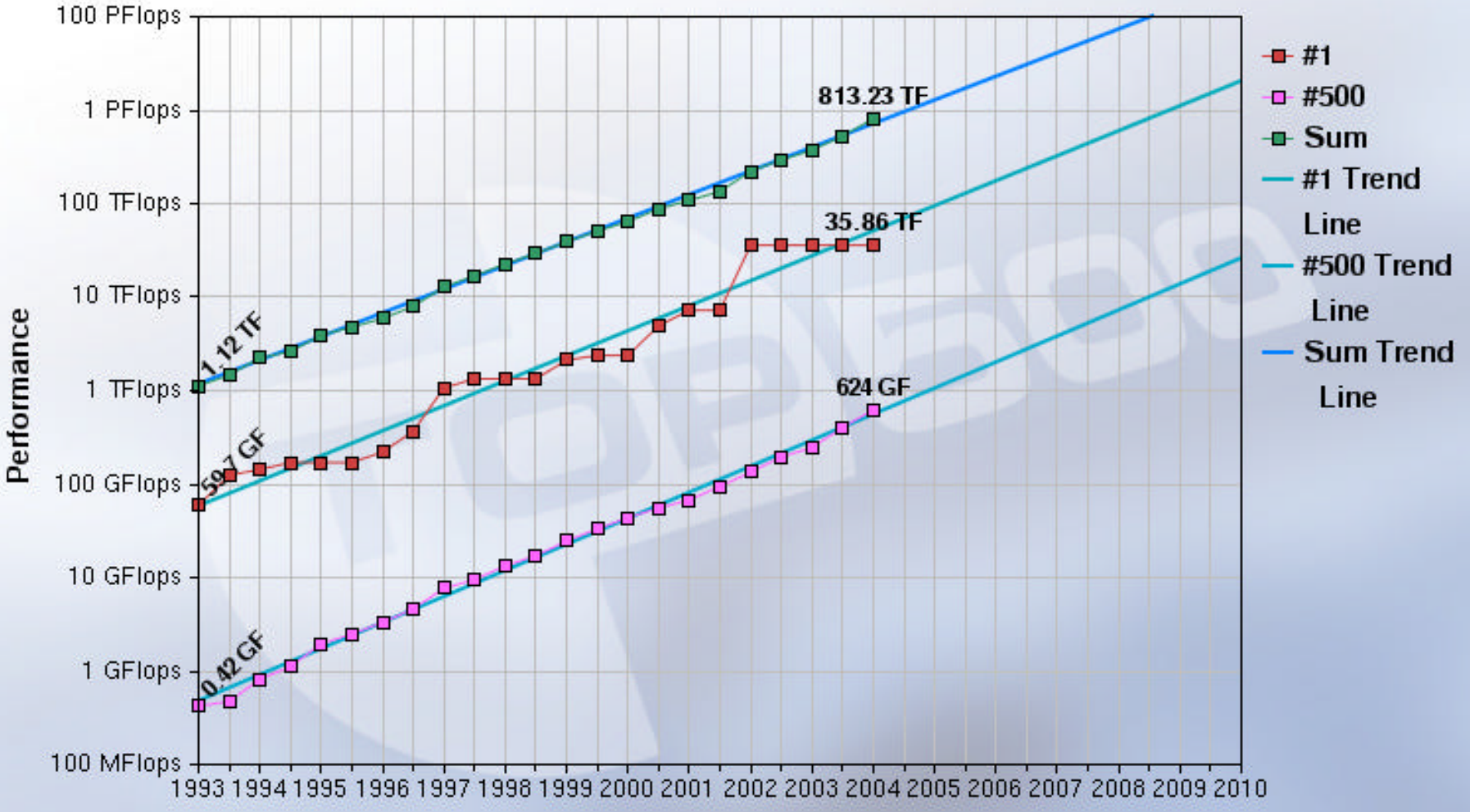
R_{max} and R_{peak} values are in GFlops. For more details about other fields, please click on the button "Explanation of the Fields"

DETAILS EXPLANATION OF THE FIELDS

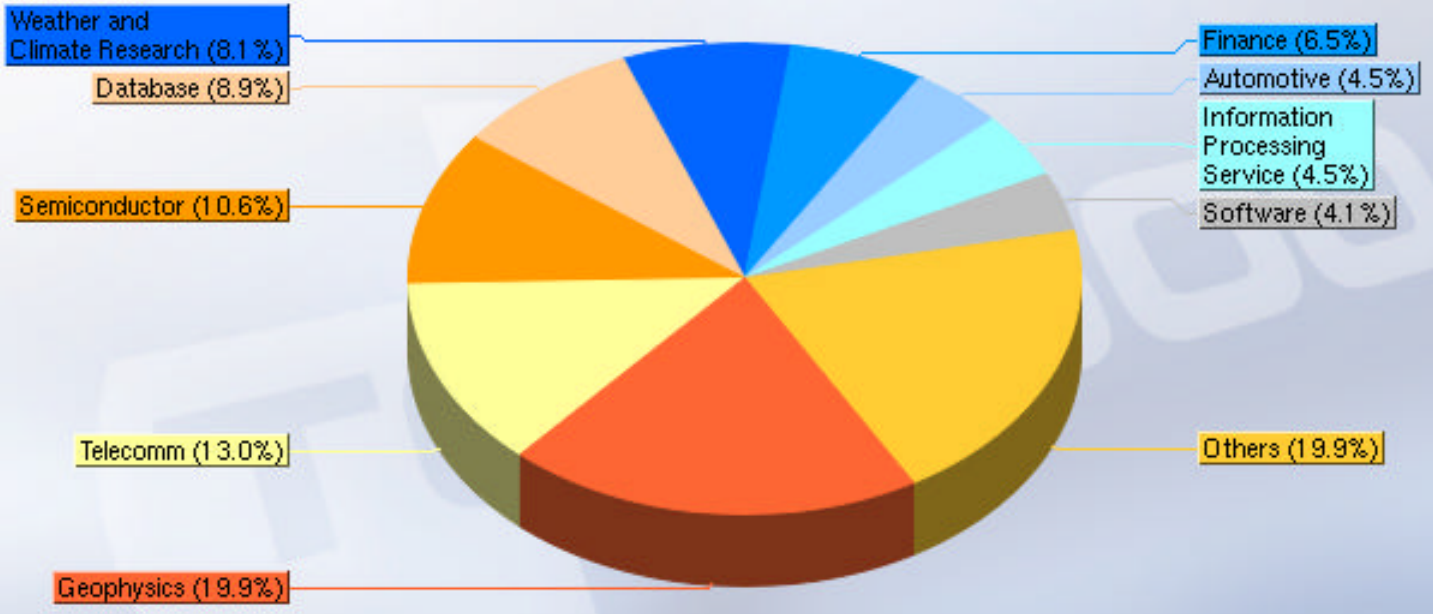
1-100 | 101-200 | 201-300 | 301-400 | 401-500

Rank	Site Country/Year	Computer / Processors Manufacturer	R_{max} R_{peak}
1	Earth Simulator Center Japan/2002	**Earth-Simulator** / 5120 NEC	35860 40960
2	Lawrence Livermore National Laboratory United States/2004	*Thunder* **Intel Itanium2 Tiger4 1.4GHz - Quadrics** / 4096 California Digital Corporation	19940 22938
3	Los Alamos National Laboratory United States/2002	**ASCI Q - AlphaServer SC45, 1.25 GHz** / 8192 HP	13880 20480
4	IBM - Rochester United States/2004	**BlueGene/L DD1 Prototype (0.5GHz PowerPC 440 w/Custom)** / 8192 IBM/ LLNL	11680 16384
5	NCSA United States/2003	*Tungsten* **PowerEdge 1750, P4 Xeon 3.06 GHz, Myrinet** / 2500 Dell	9819 15300
6	ECMWF United Kingdom/2004	**eServer pSeries 690 (1.9 GHz Power4+)** / 2112 IBM	8955 16051
7	Institute of Physical and Chemical Res. (RIKEN) Japan/2004	**RIKEN Super Combined Cluster** / 2048 Fujitsu	8728 12534

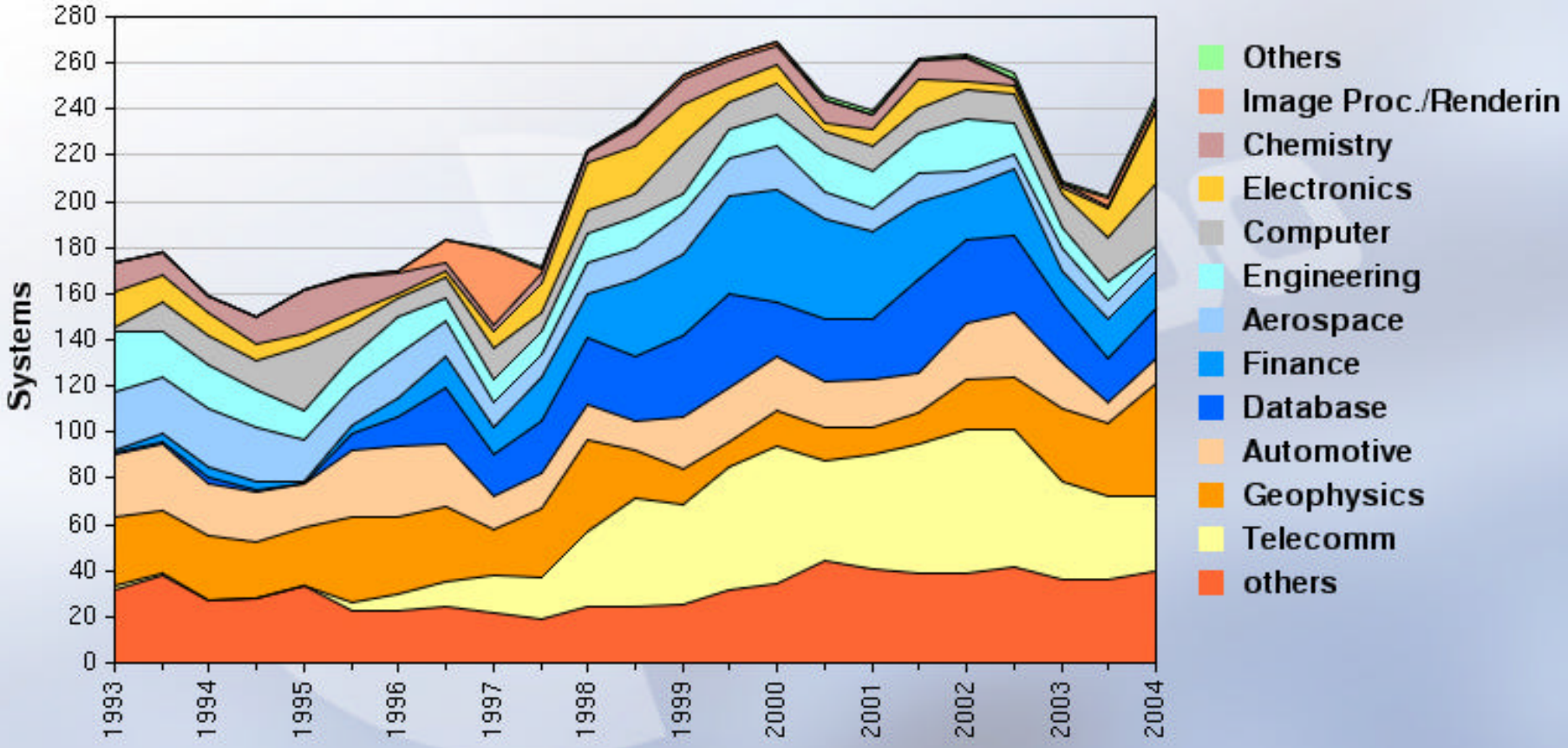
Projected Performance Development



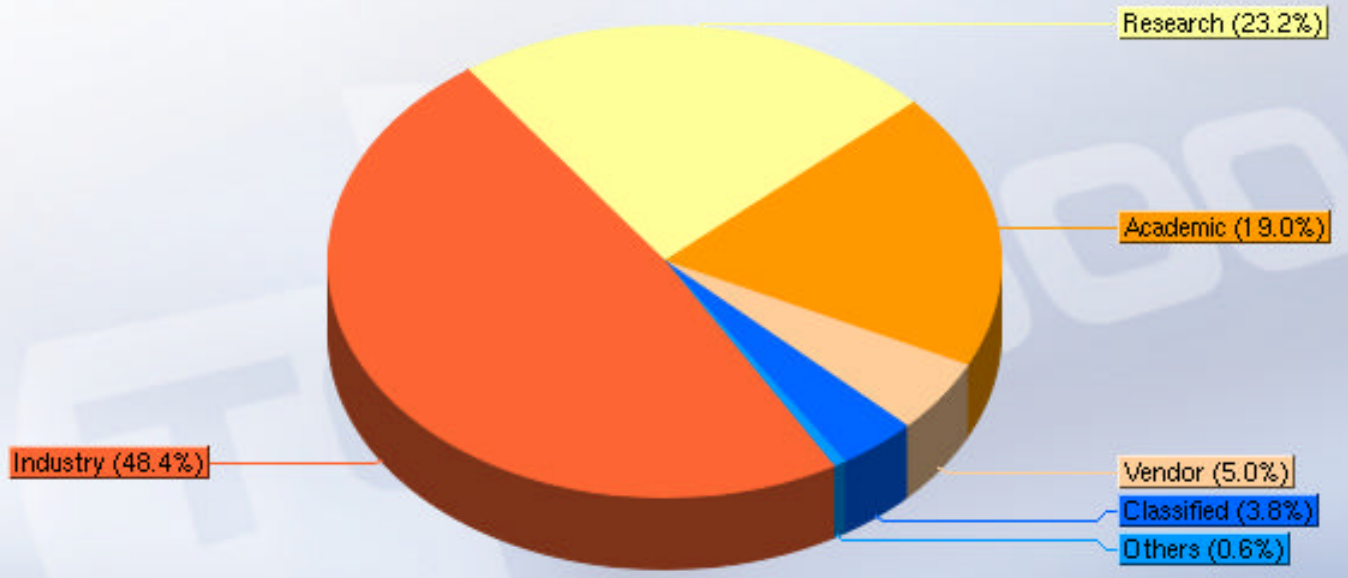
Application Area / Systems (Jun 2004)



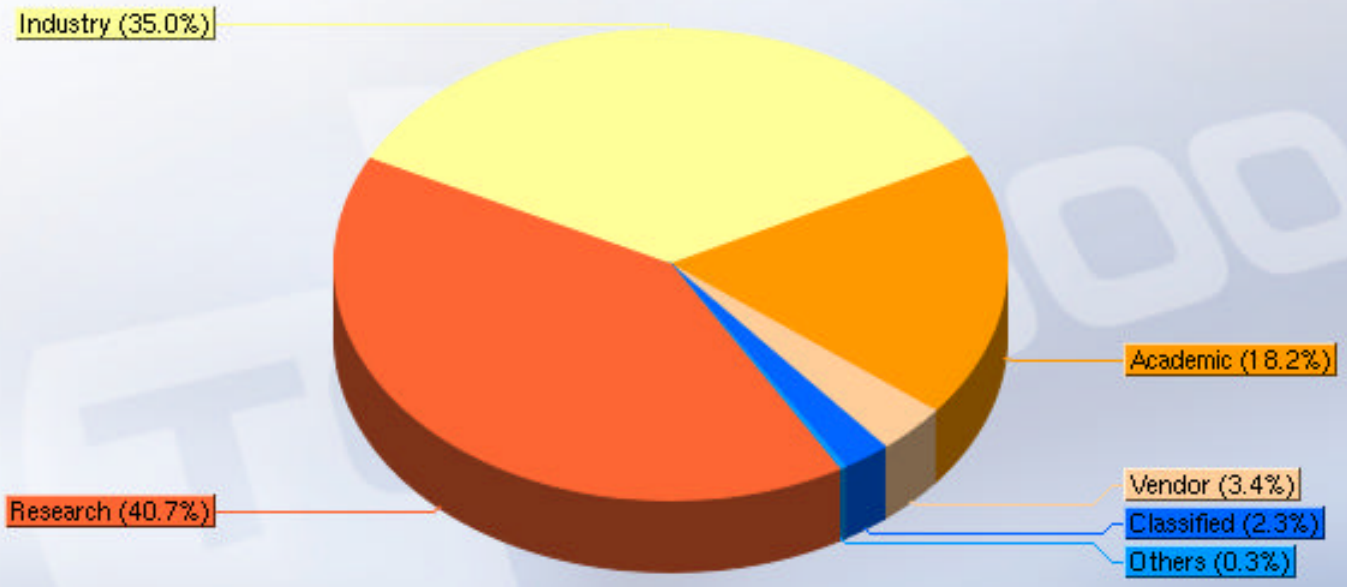
Application Area / Systems



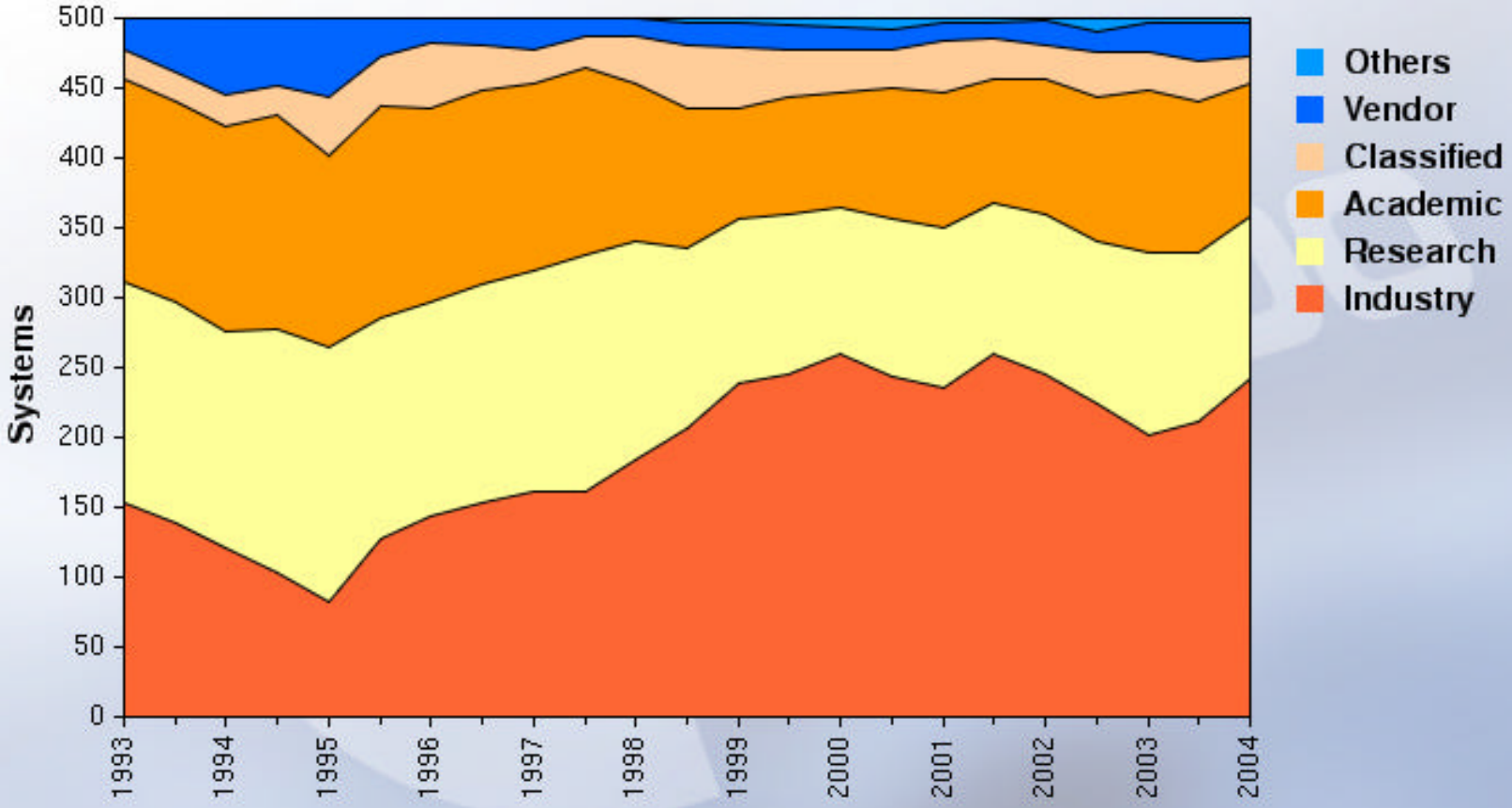
Customer Segment / Systems (Jun 2004)



Customer Segment / Performance (Jun 2004)

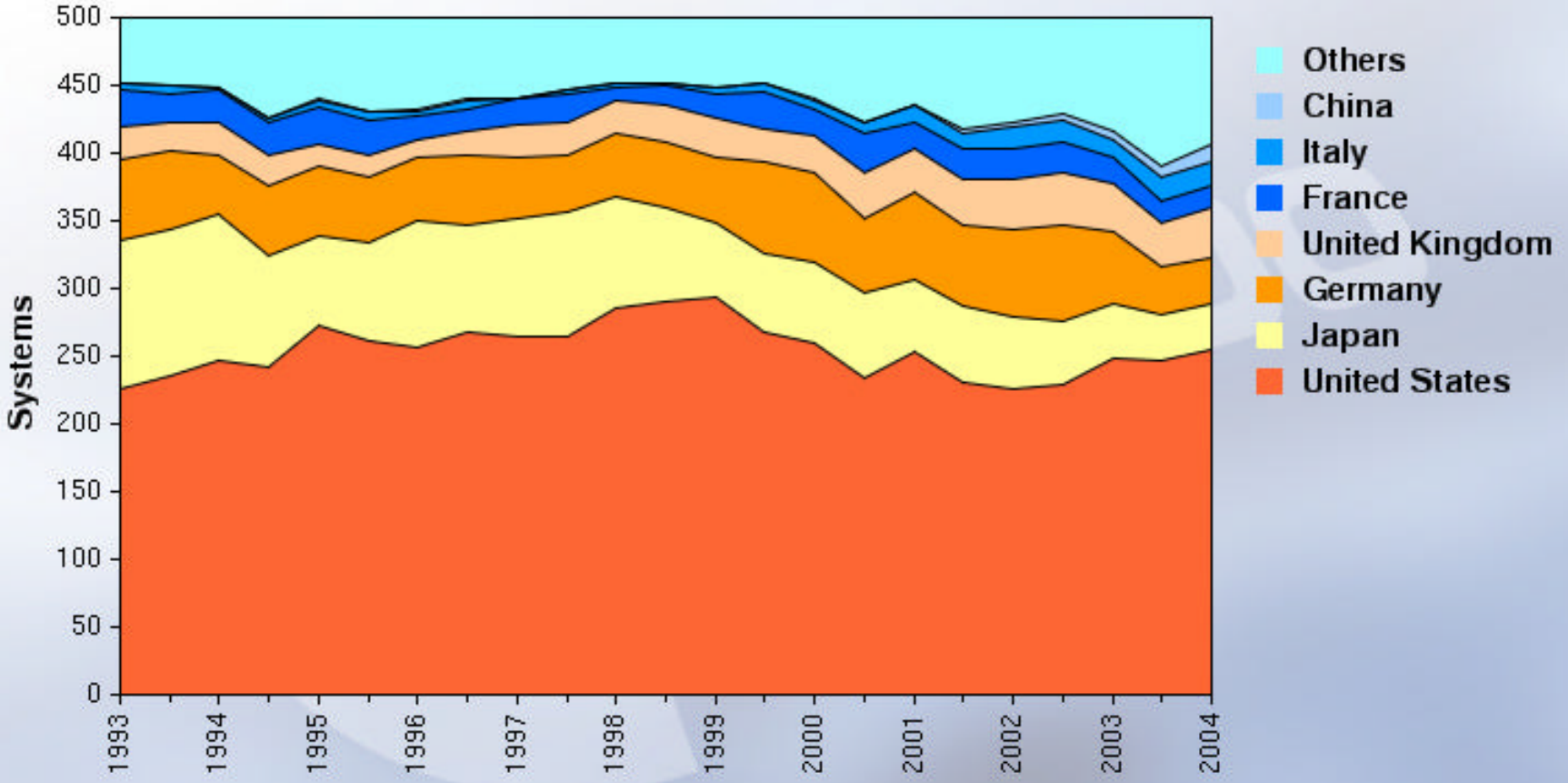


Customer Segment / Systems



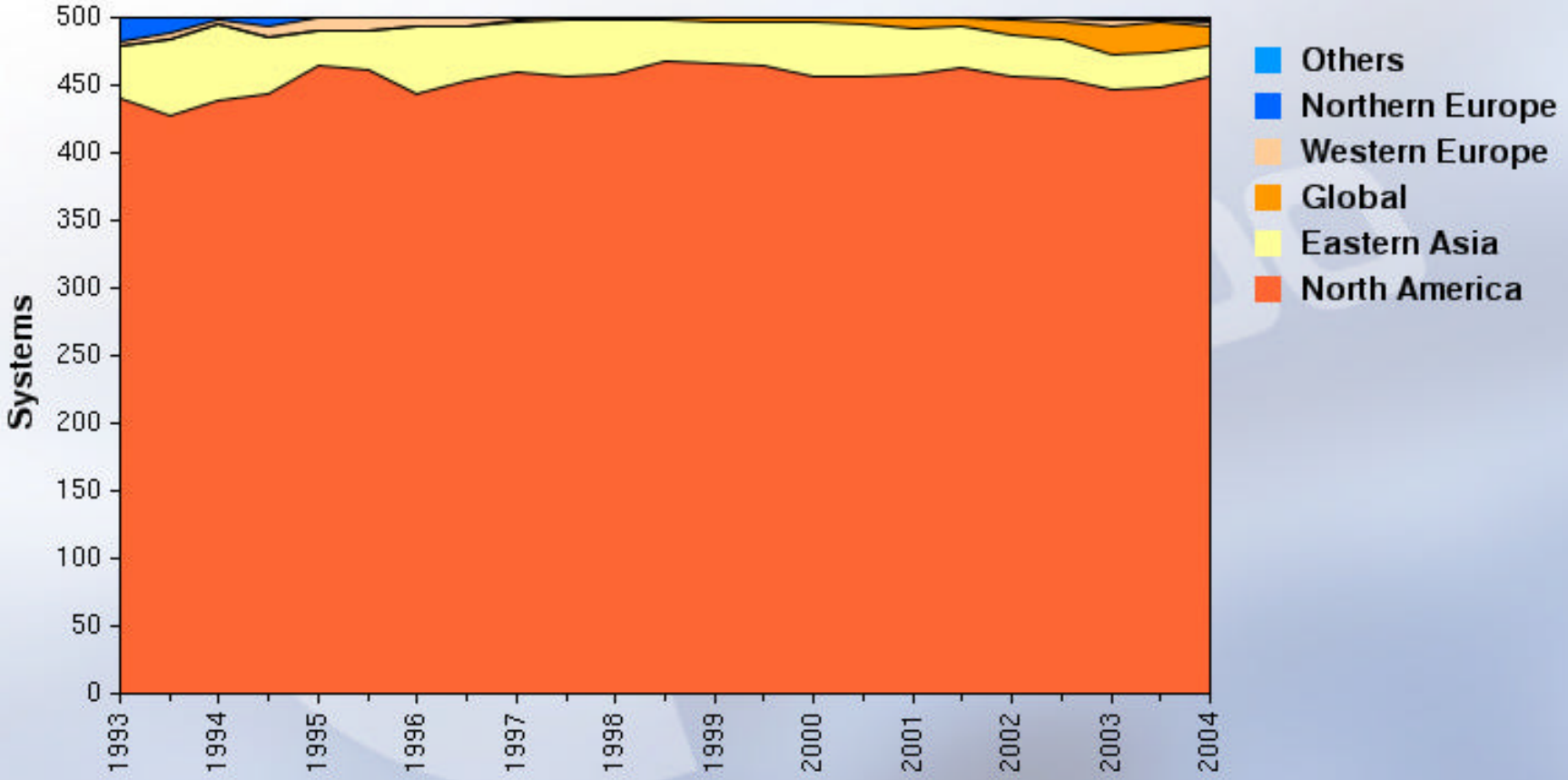
TOP500 SUPERCOMPUTER SITES

Countries / Systems



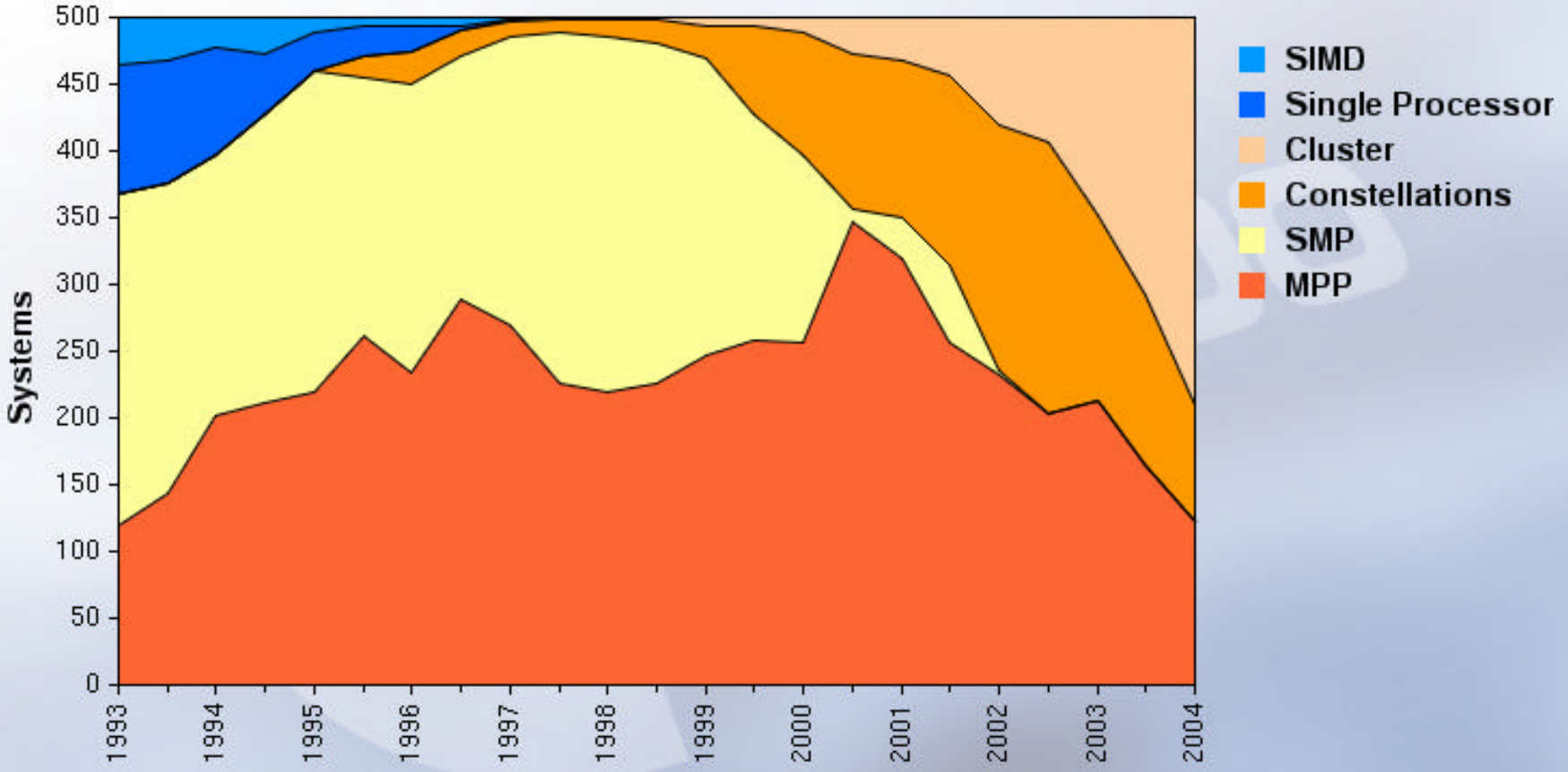
TOP500 SUPERCOMPUTER SITES

Producing Region / Systems

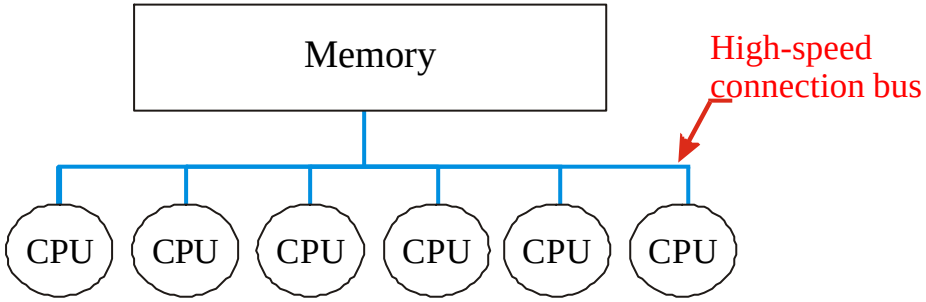


TOP500 SUPERCOMPUTER SITES

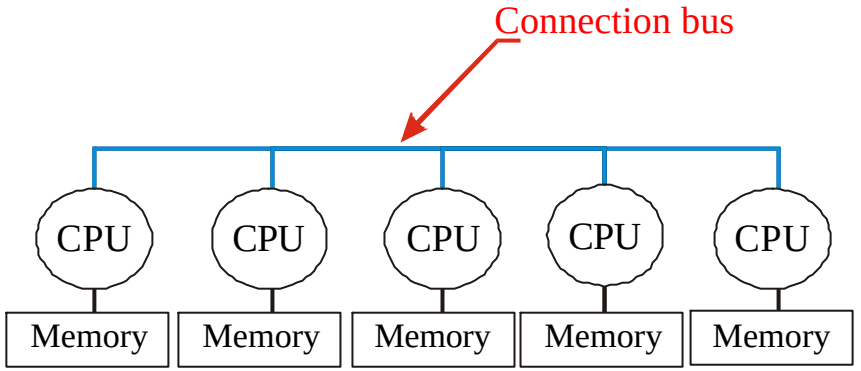
Architectures / Systems



Memória Compartilhada

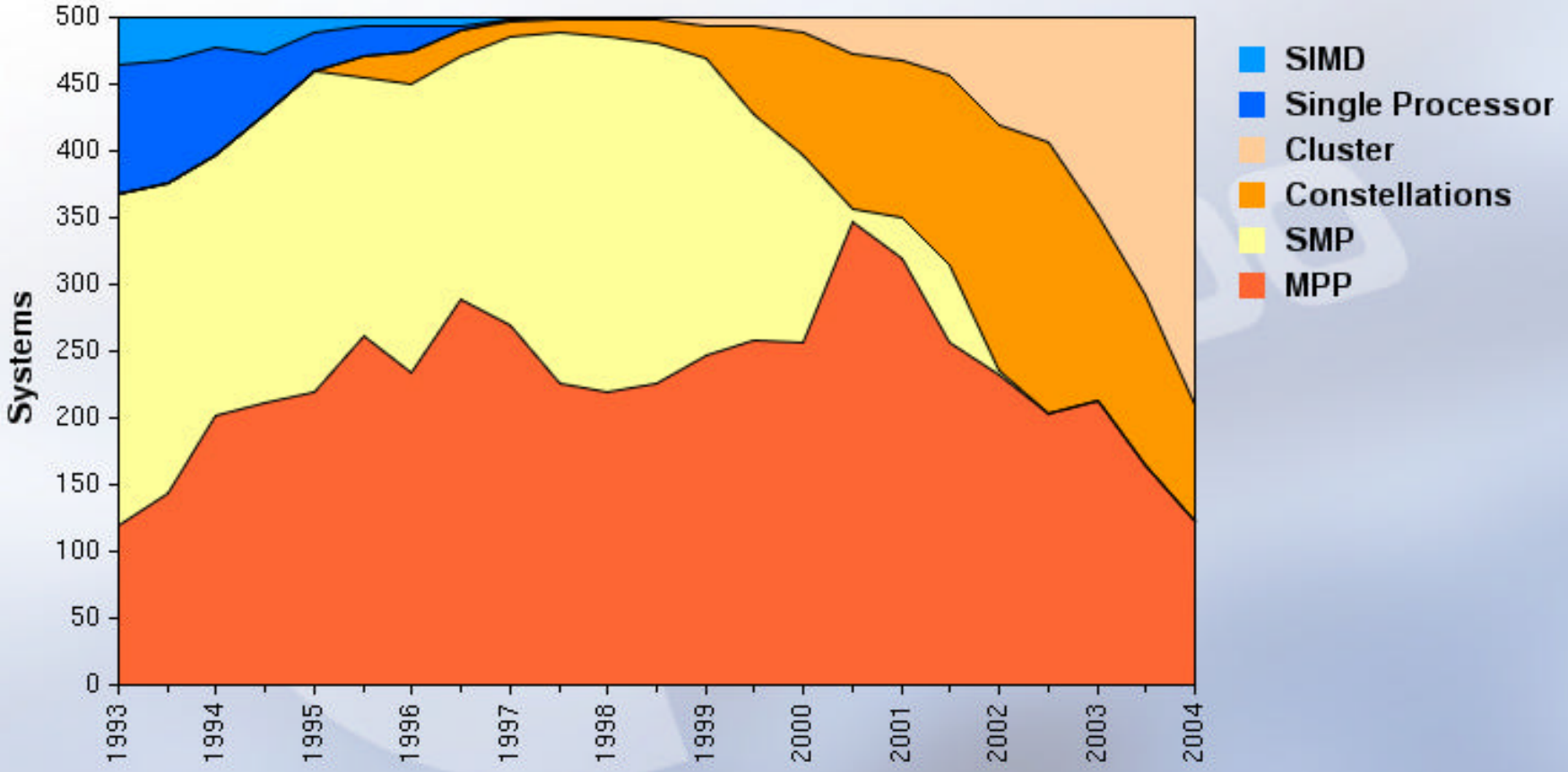


Memória Distribuída



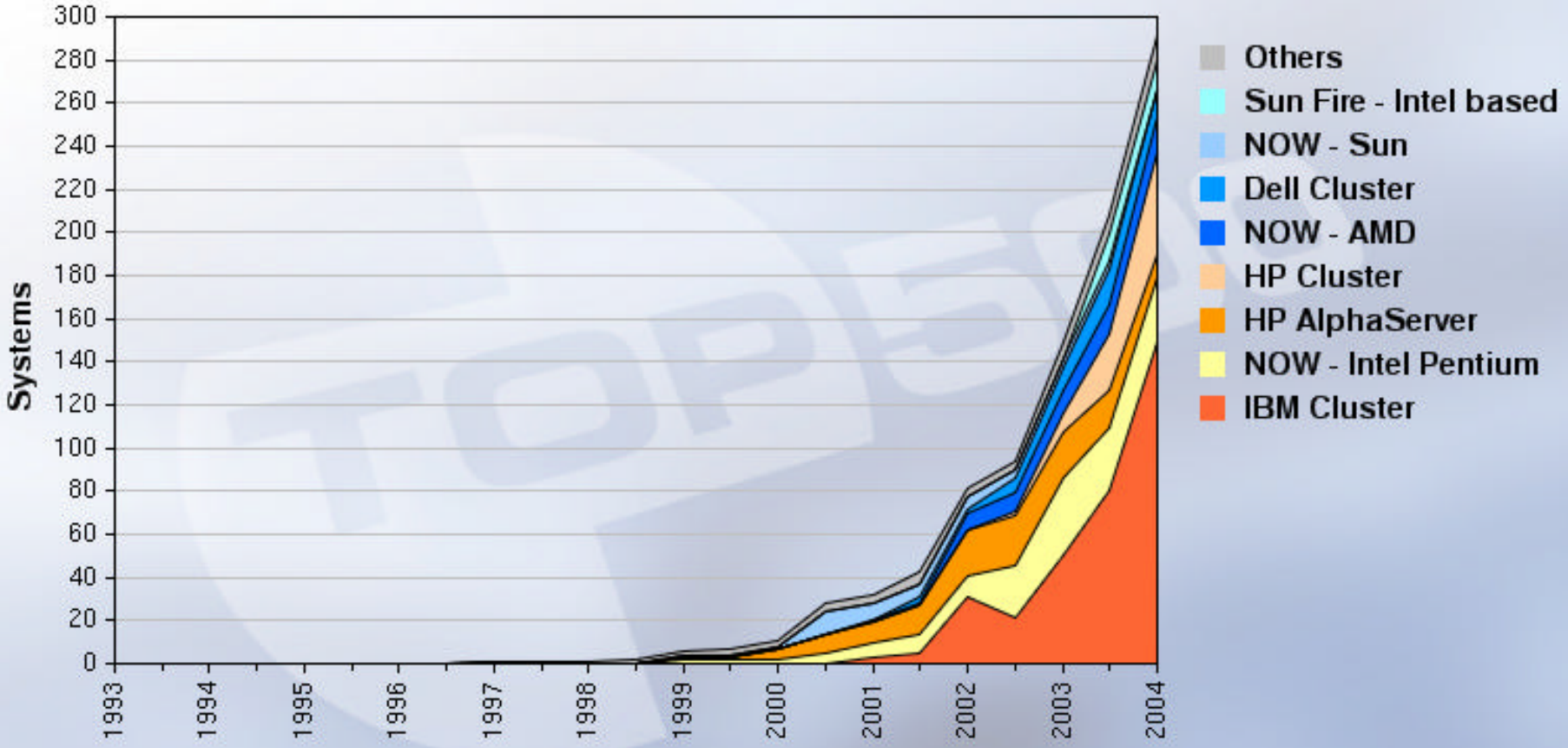
TOP500 SUPERCOMPUTER SITES

Architectures / Systems

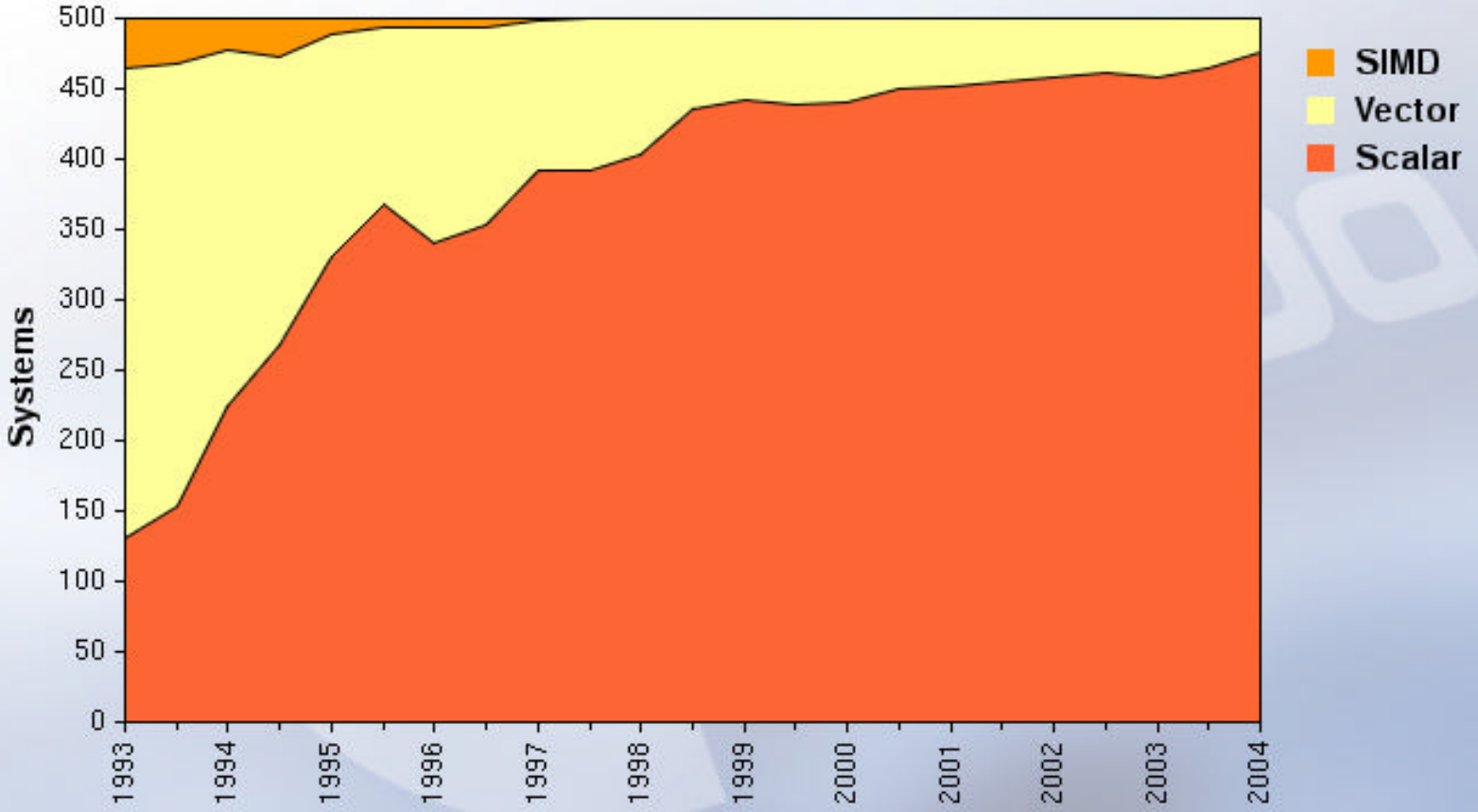


TOP500 SUPERCOMPUTER SITES

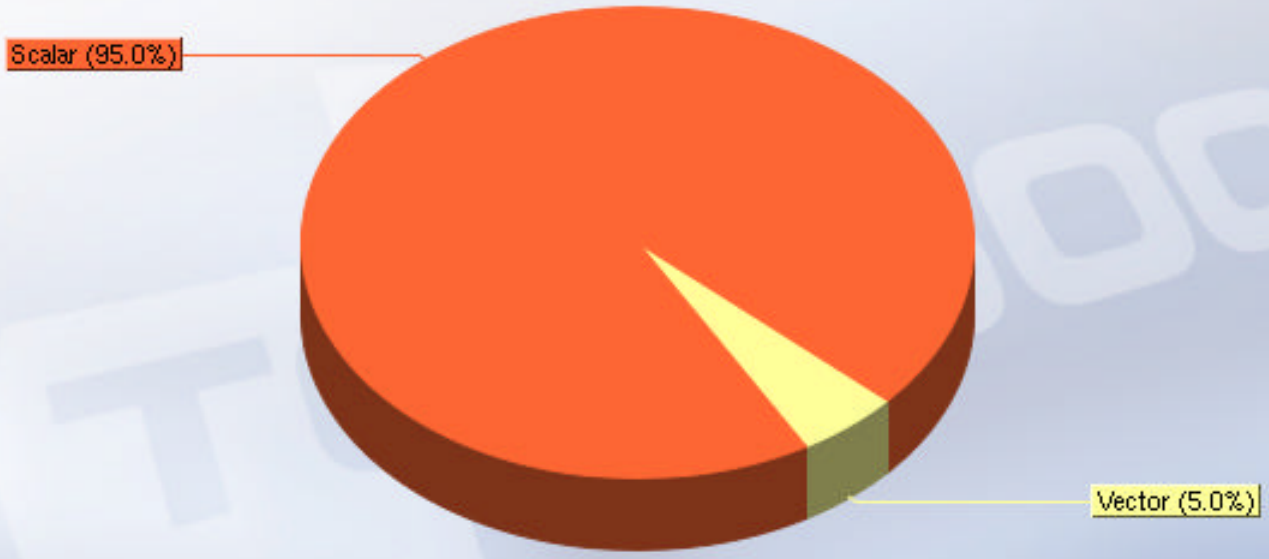
Clusters (NOW) / Systems



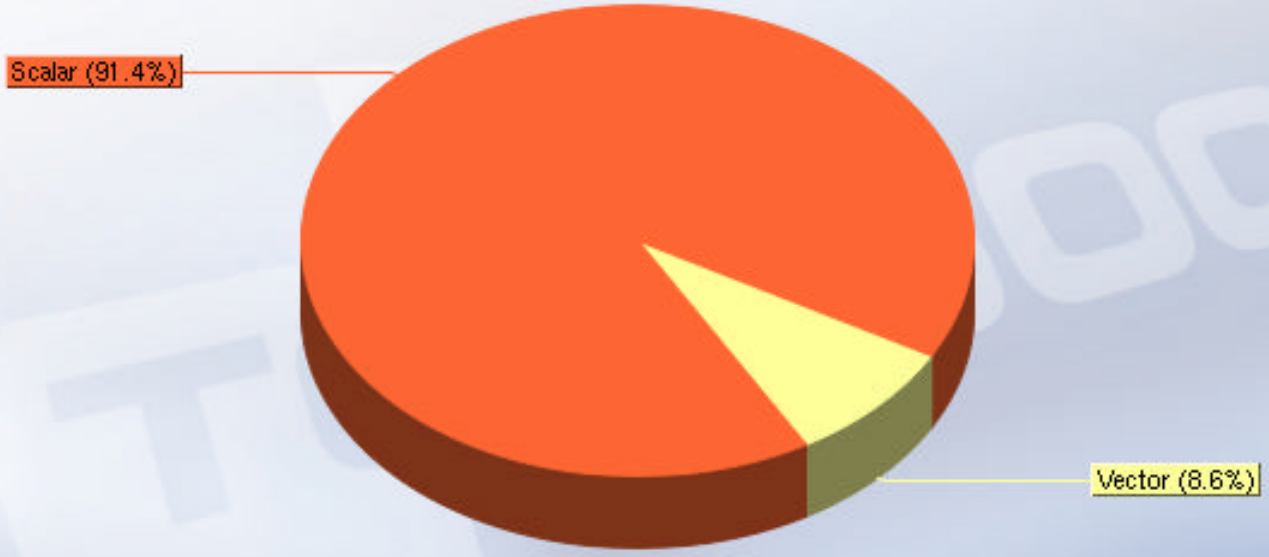
Processor Architectures / Systems



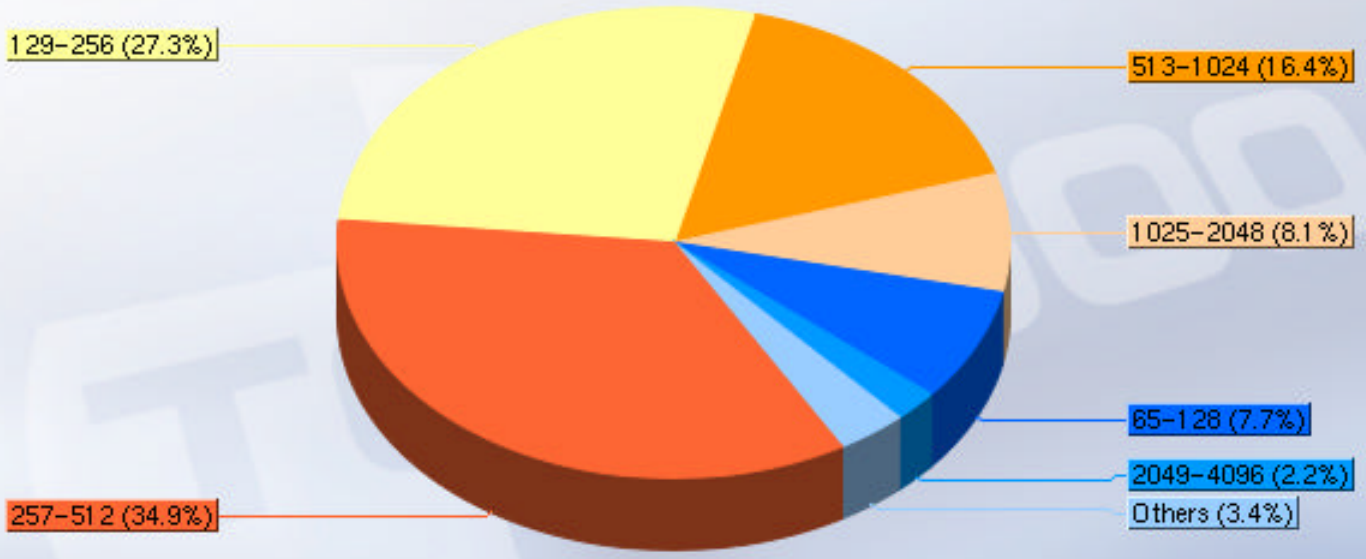
Processor Architecture / Systems (Jun 2004)



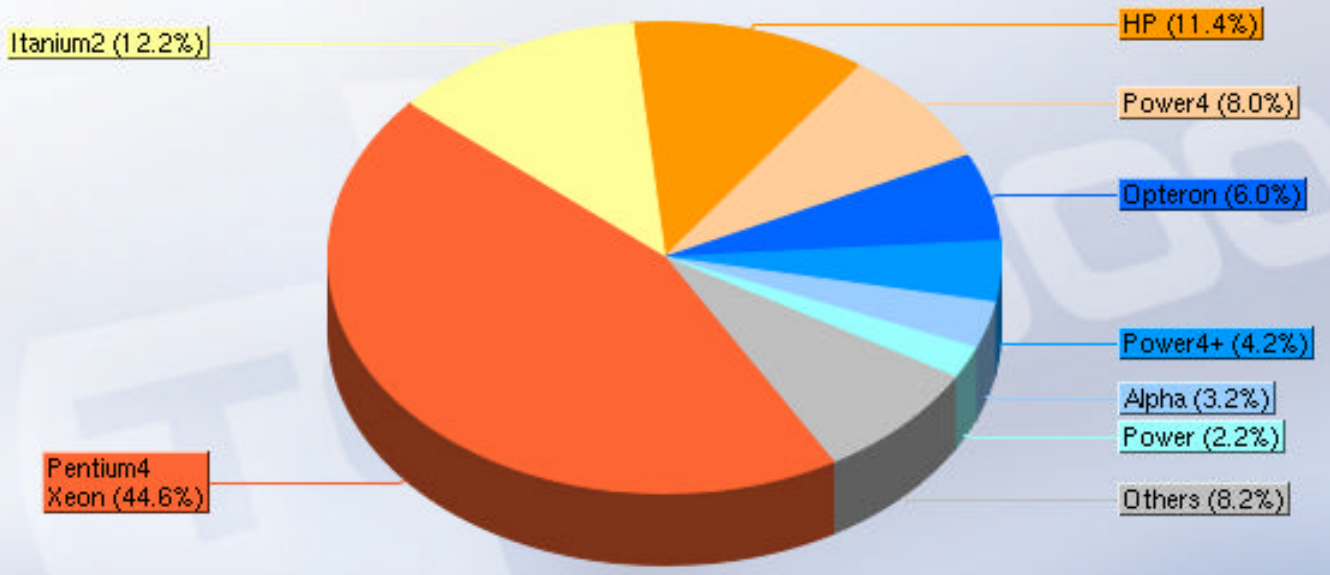
Processor Architecture / Performance (Jun 2004)



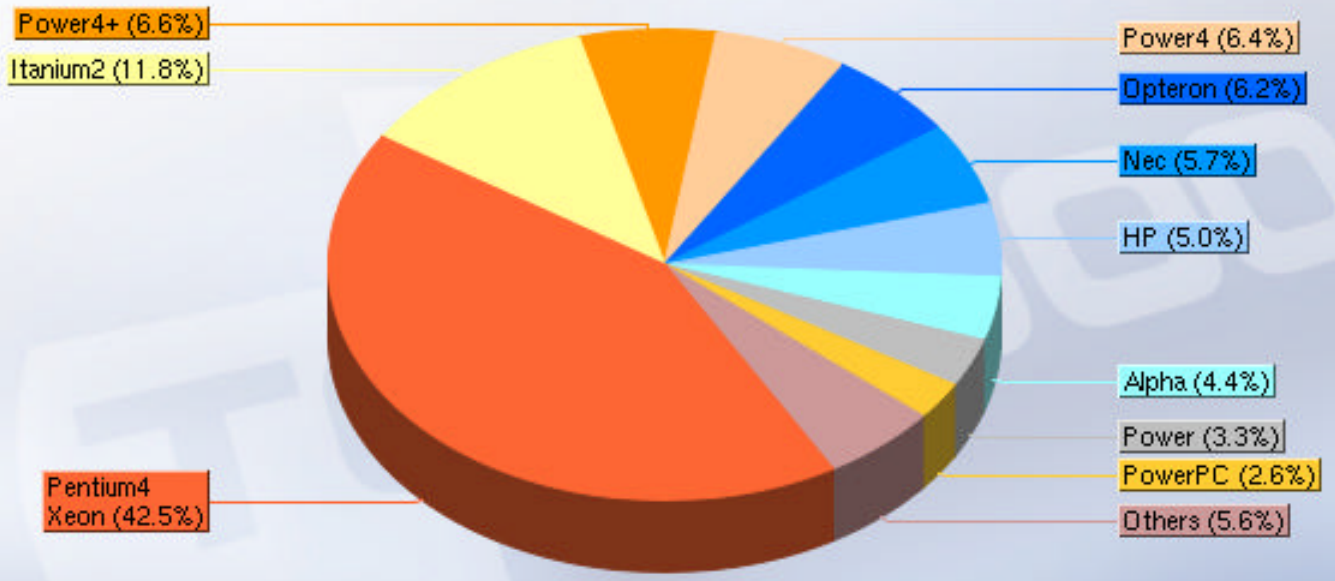
System Processor Counts / Systems (Jun 2004)



Processor Generation / Systems (Jun 2004)

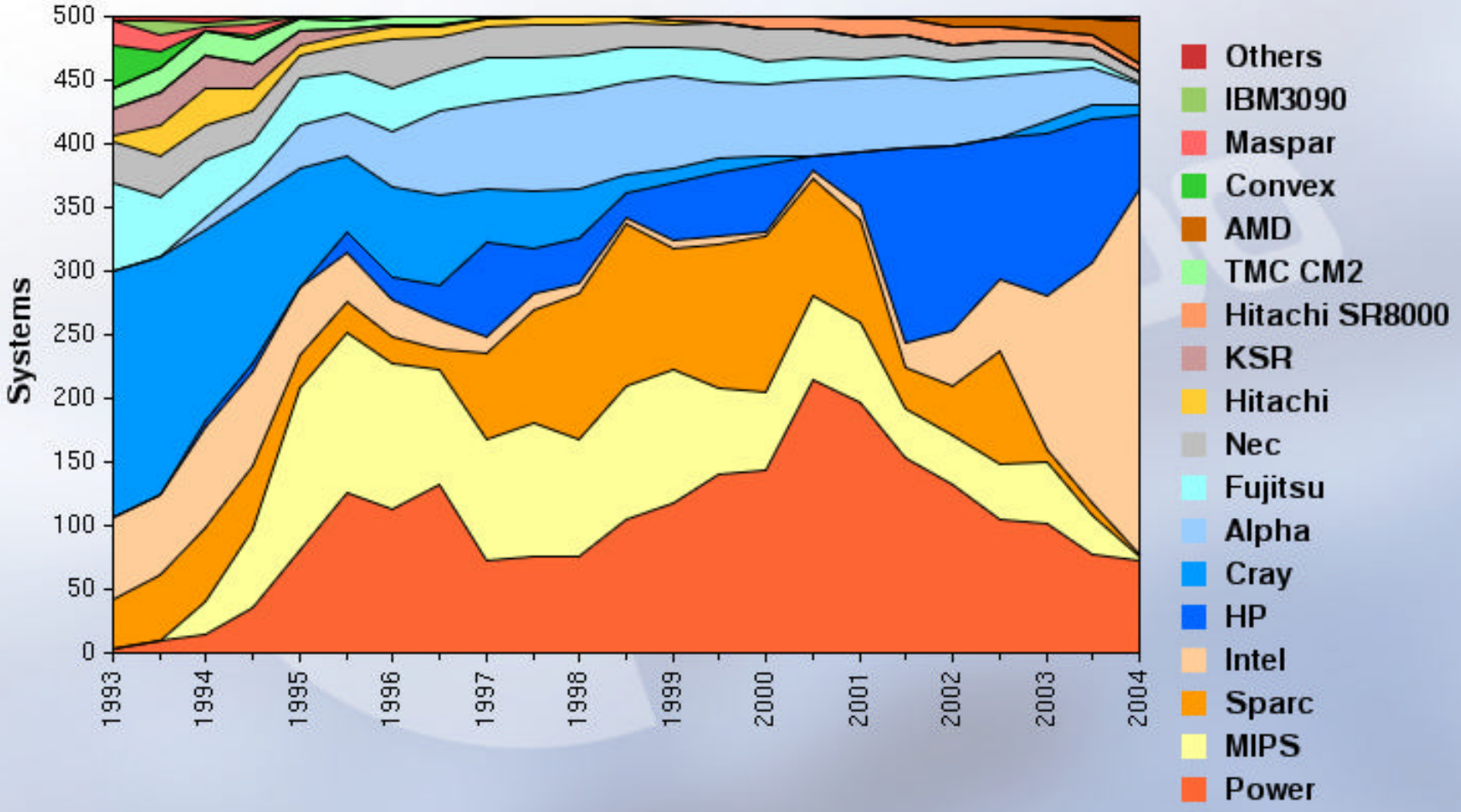


Processor Generation / Performance (Jun 2004)

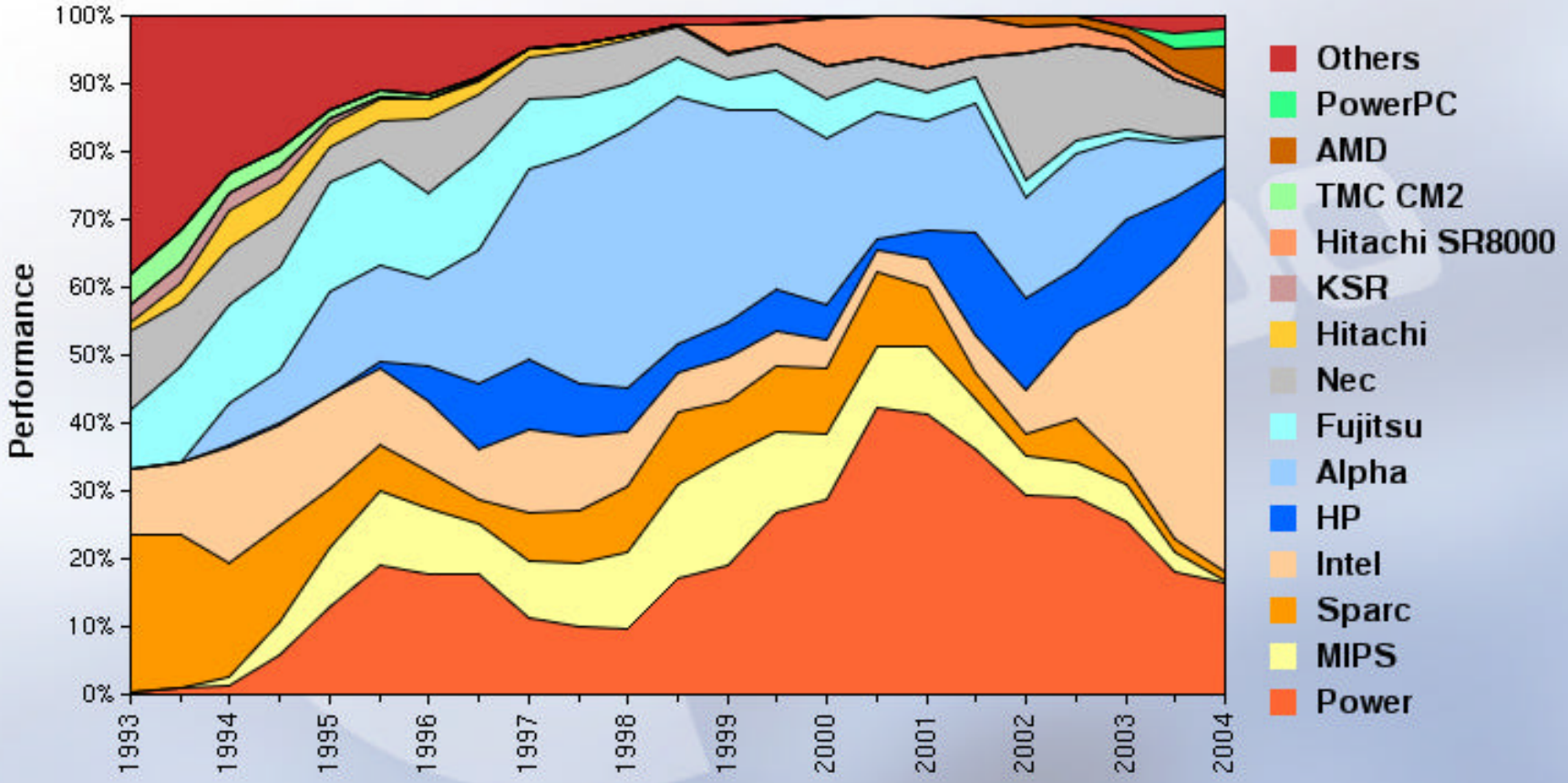


TOP500 SUPERCOMPUTER SITES

Processor Family / Systems

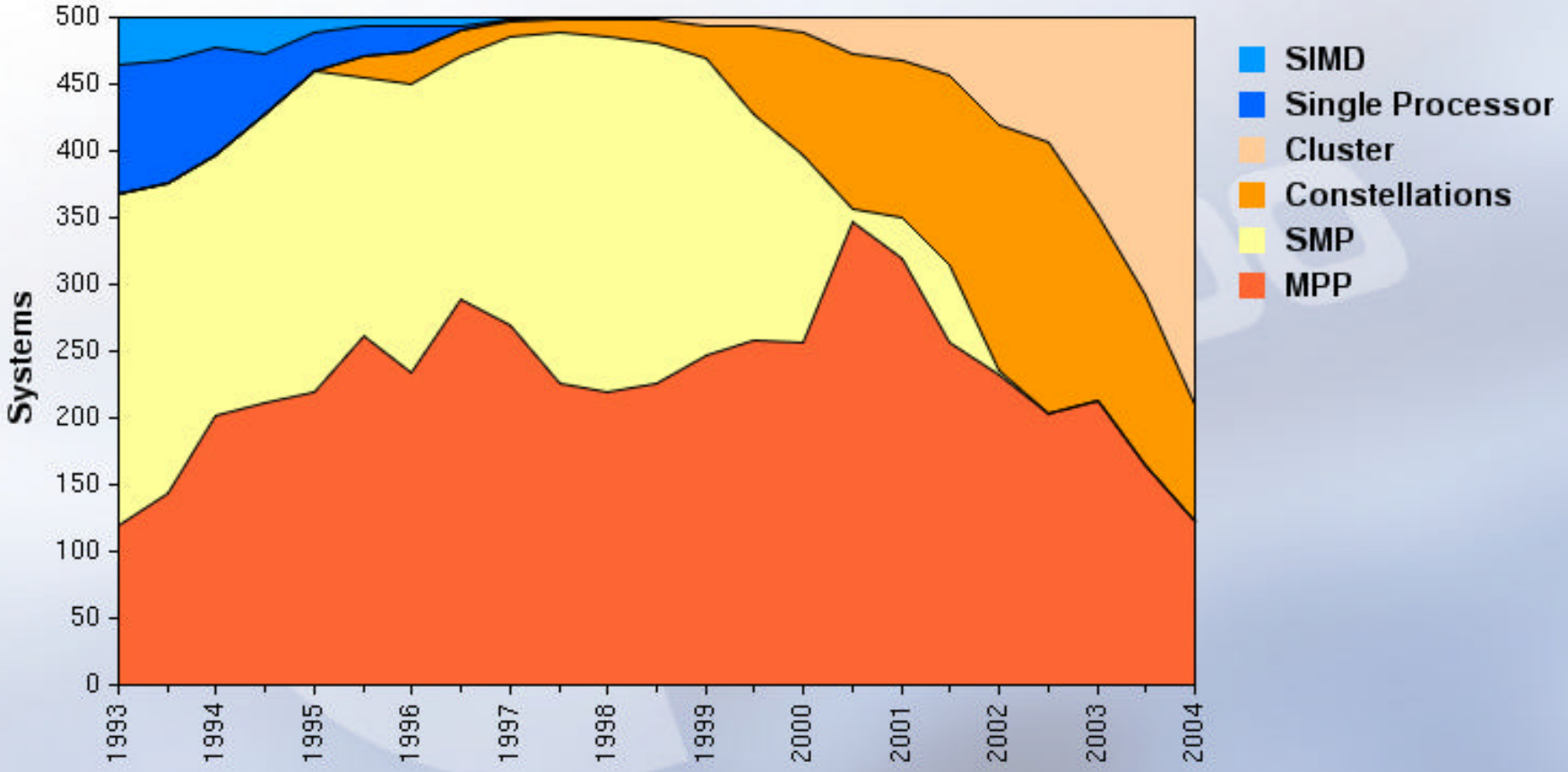


Processor Family / Performance



TOP500 SUPERCOMPUTER SITES

Architectures / Systems



TOP 5 SUPERCOMPUTER SITES (June 2004)

1

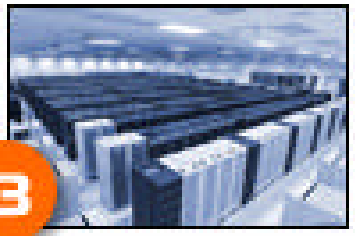
EARTH SIMULATOR

Earth Simulator Center
Yokohama, Japan
NEC SX6
Rmax: 35.86 TFlops

2

Thunder

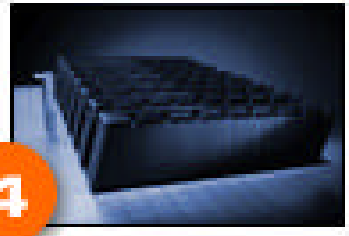
LLNL
Livermore, USA
Intel Tiger4/Quadrics
Rmax: 19.94 TFlops



3

ASCI Q

LANL
Los Alamos, USA
HP Alphaserver SC
Rmax: 13.88 TFlops



4

BlueGene

Thomas Watson Research Center
Yorktown Heights, USA
BlueGene/L DD1 Prototype
Rmax: 11.68 TFlops

5

Tungsten

NCSA
Urbana-Champaign, USA
Dell PowerEdge 1750/Myrinet
Rmax: 9.819 TFlops

1

EARTH SIMULATOR

EARTH SIMULATOR CENTER
YOKOHAMA
NEC SX
Rmax: 35.86 TFlops

Introduction

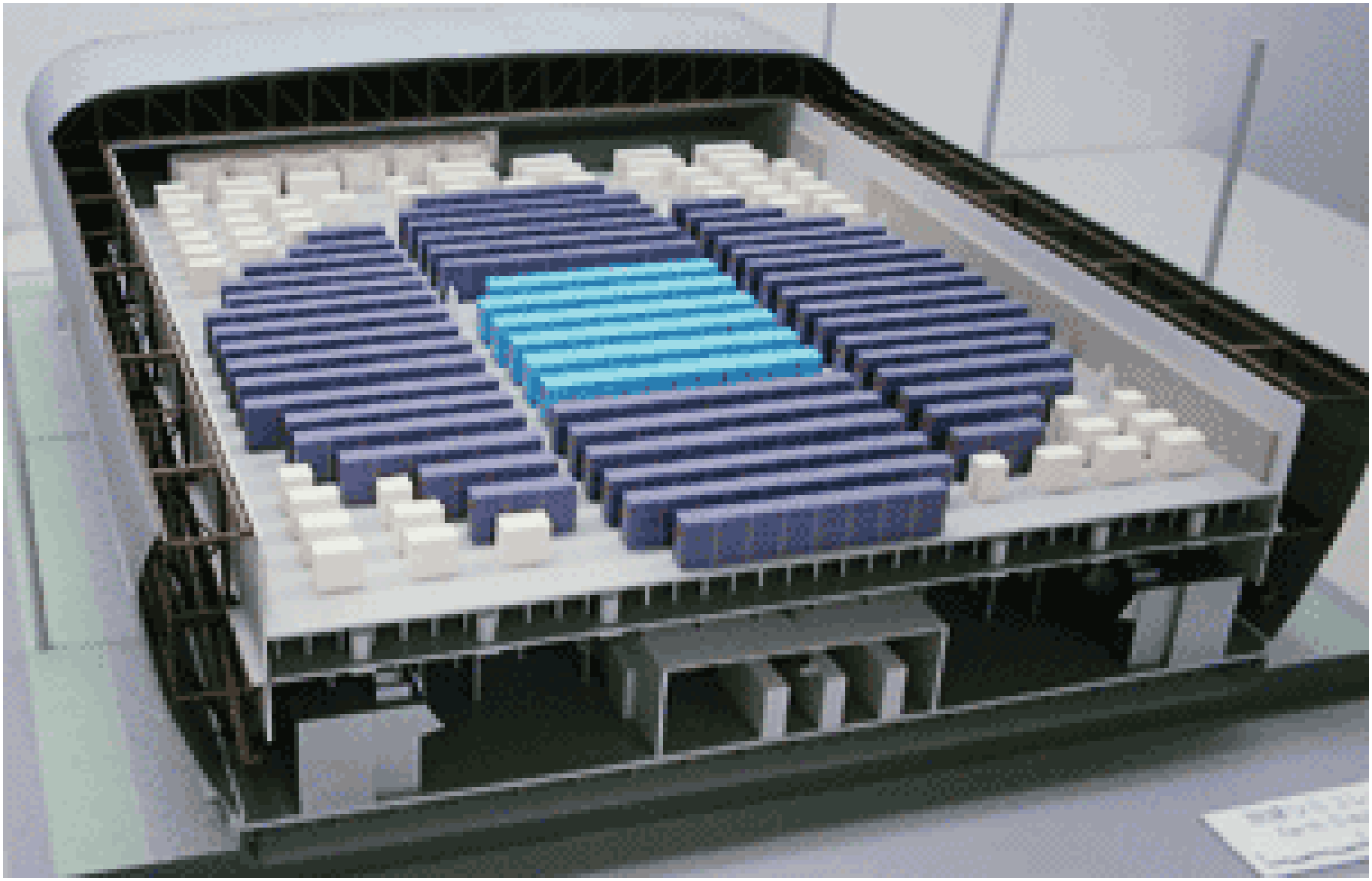
The Earth Simulator (ES) is a project of Japanese agencies NASDA, JAERI and JAMSTEC to develop a 40 TFLOPS system for climate modeling.

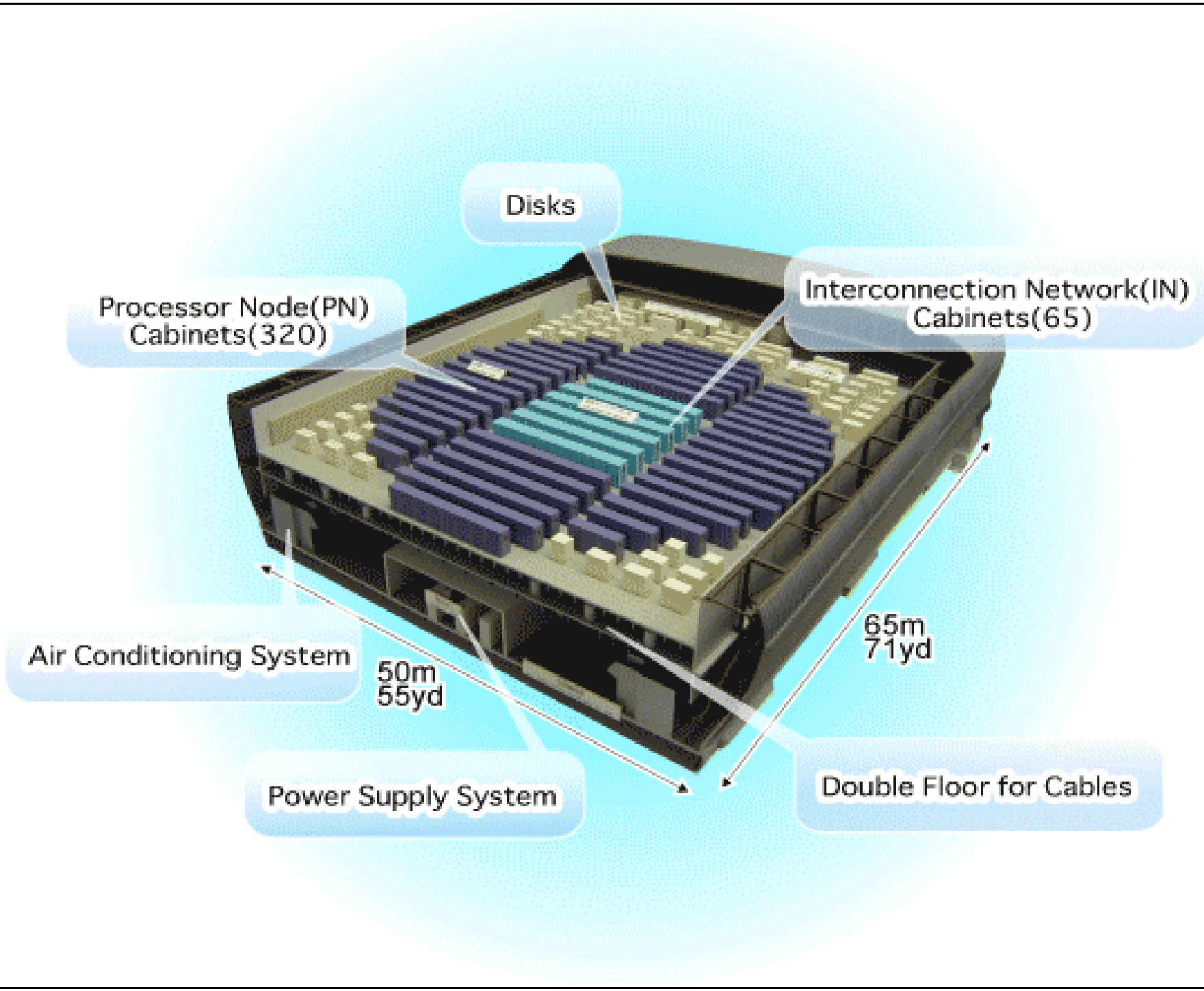
The ES site is a new location in an industrial area of Yokohama, an hour drive west of Tokyo. The facility became operational in late 2001, and claimed first spot in current Top 500 list. In spite of a public 5-year development period leading to that event, the US supercomputing community was caught by surprise.

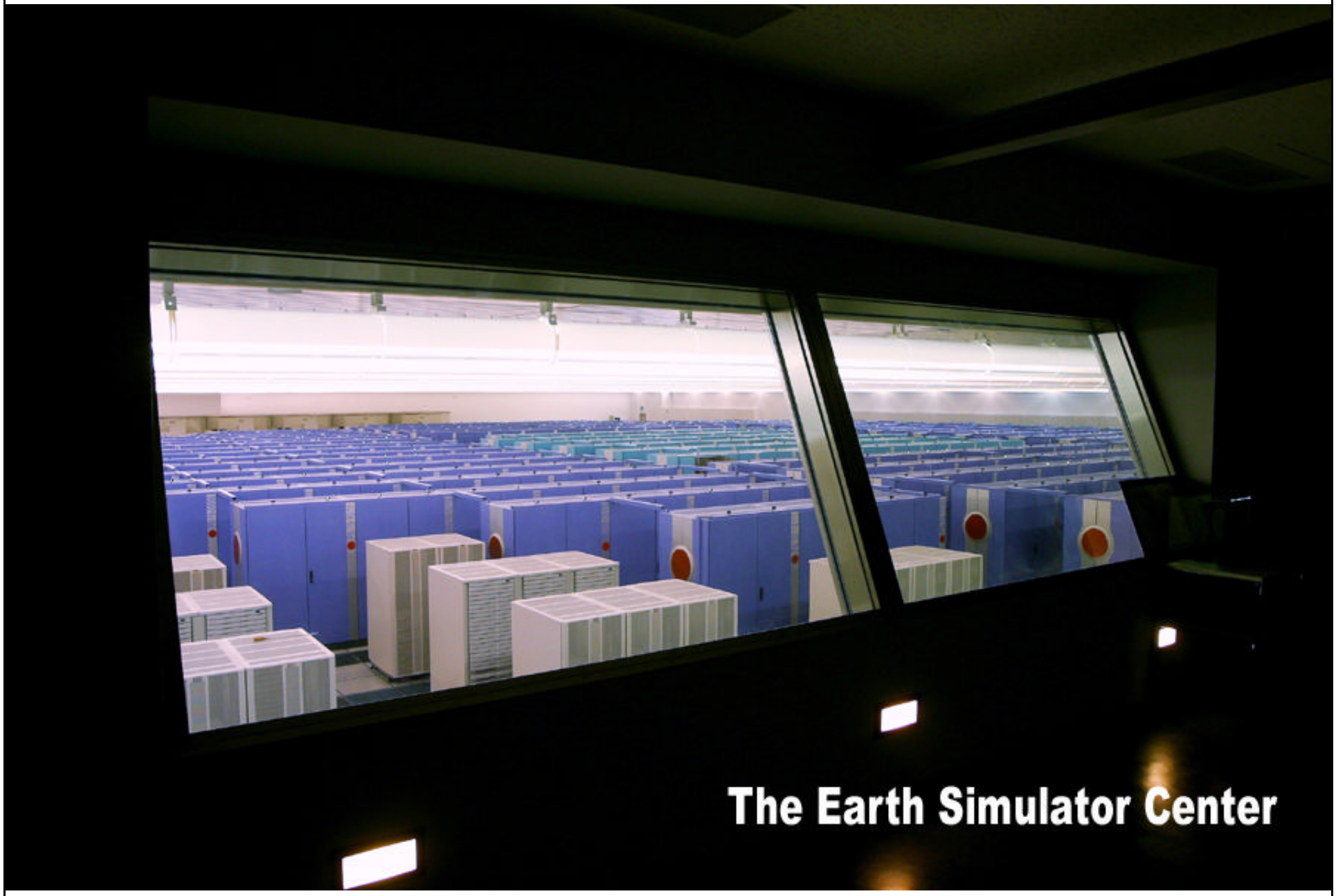
Hardware

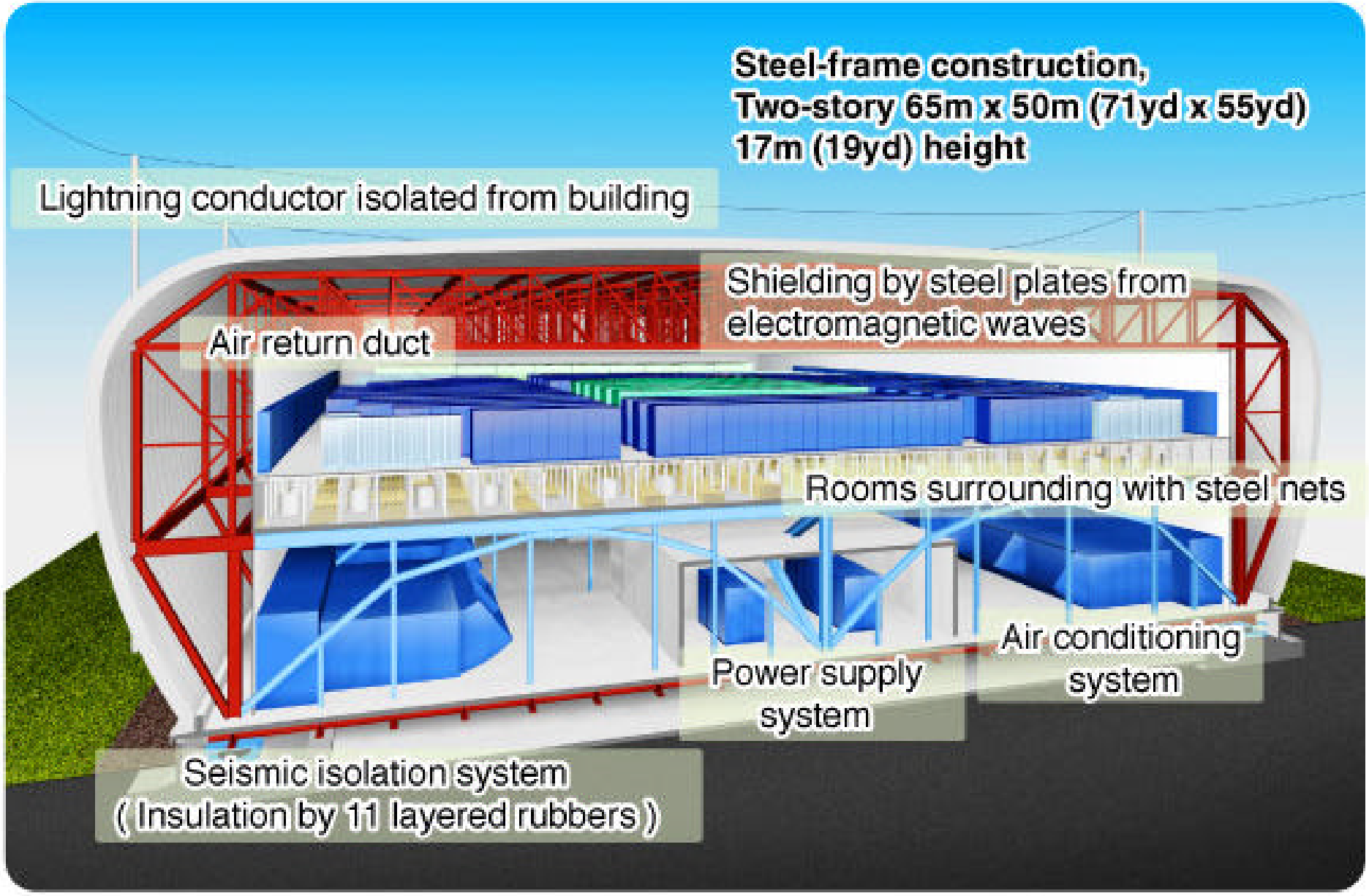
The ES is based on:

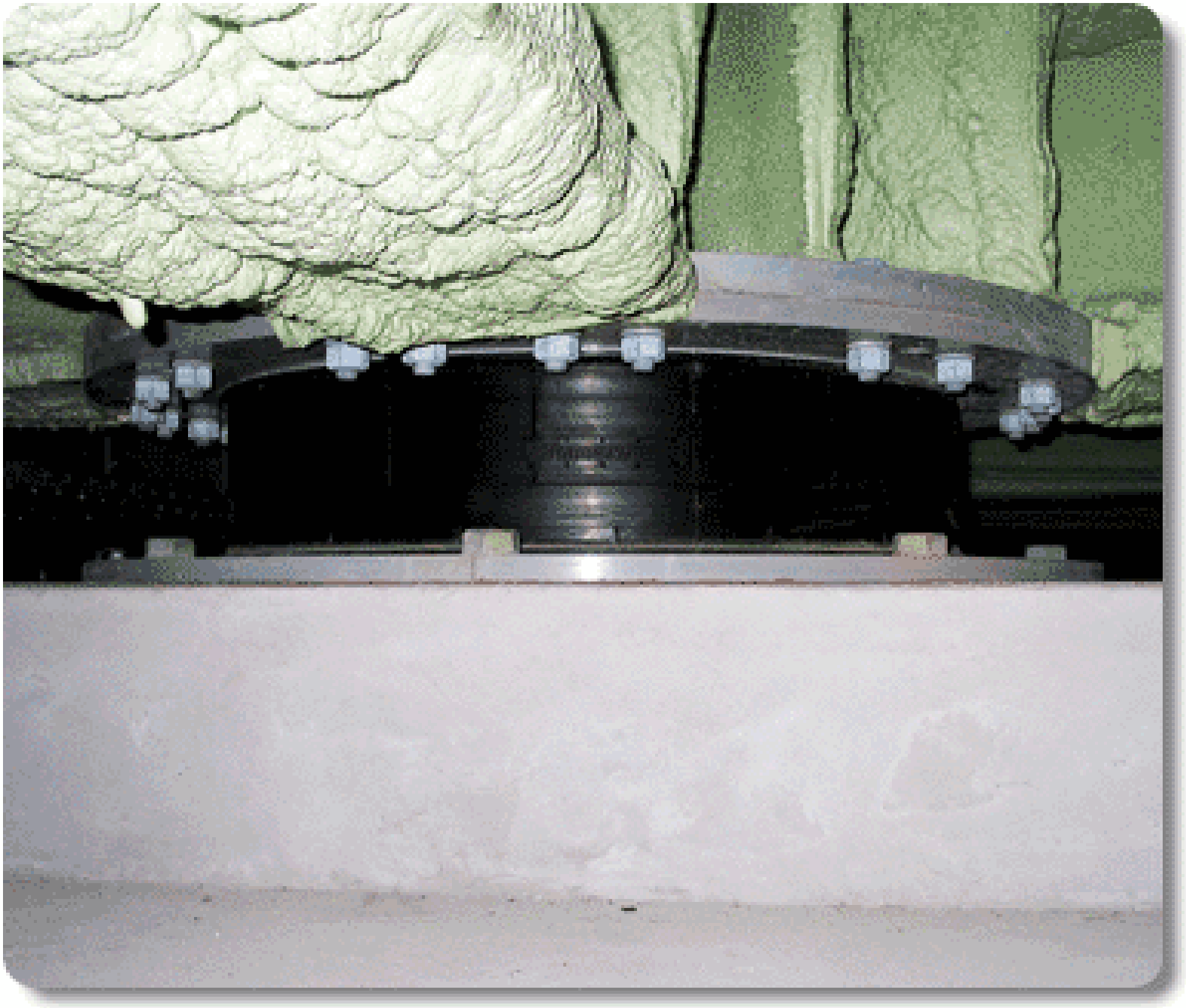
- 5,120 (640 8-way nodes) 500 MHz NEC CPUs
- 8 GFLOPS per CPU (41 TFLOPS total)
- 2 GB (4 512 MB FPLRAM modules) per CPU (10 TB total)
- shared memory inside the node




Disks
Processor Node(PN)
Cabinets(320)
Interconnection Network(IN)
Cabinets(65)
Air Conditioning System
50m
55yd
65m
71yd
Power Supply System
Double Floor for Cables

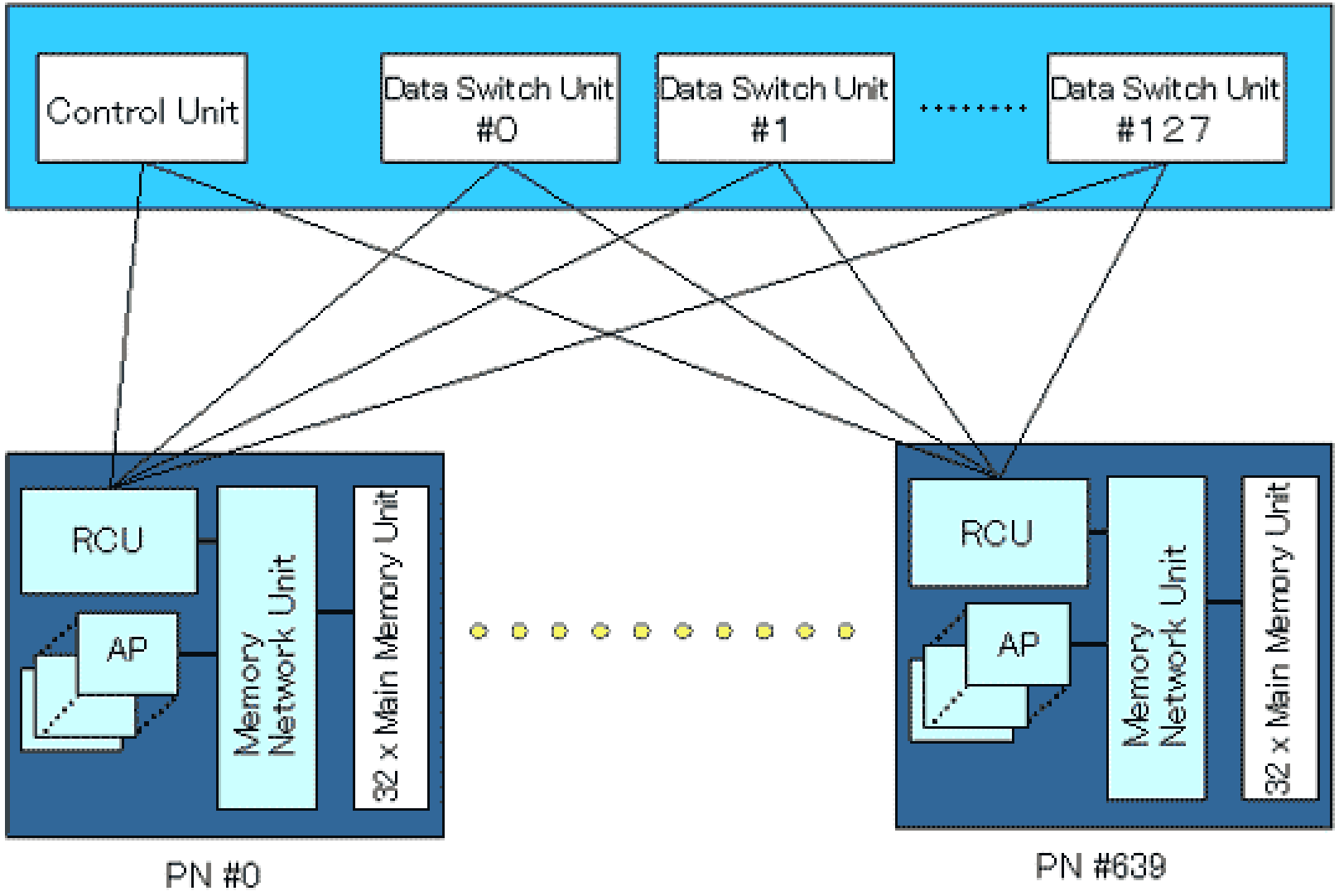

The Earth Simulator Center


Steel-frame construction,
Two-story 65m x 50m (71yd x 55yd)
17m (19yd) height
Lightning conductor isolated from building
Shielding by steel plates from electromagnetic waves
Air return duct
Rooms surrounding with steel nets
Air conditioning system
Power supply system
Seismic isolation system
(Insulation by 11 layered rubbers)

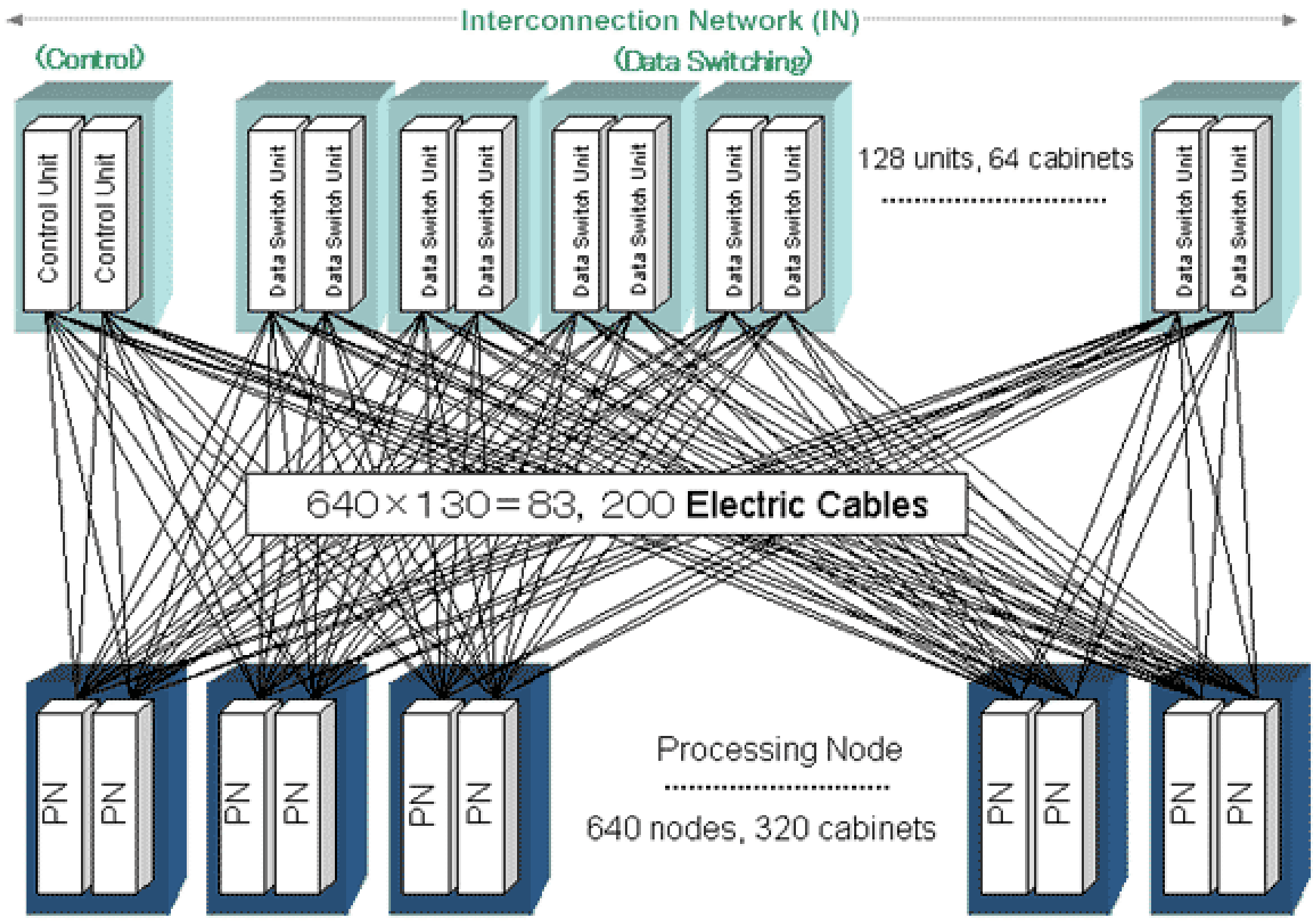


Construction of Interconnection Network

Interconnection Network



Interconnection Network



2

Thunder

LLNL
Livermore, USA
Intel Tiger4/Quadrics
Rmax: 19.94 TFlops

Thunder is a highly integrated, well-balanced capability compute resource with 1,024 nodes (see the detailed system architecture and simulation environment). The cluster will be capable of supporting a complex workload consisting of medium (1,024-2,048) and large (2,048-4,004) MPI task count parallel jobs for unclassified Multiprogrammatic & Institutional Computing (M&IC) simulations.

Nodes	1024
Processors per node (Itanium2)	4
Total number of processors	4096
Processor speed (GHz)	1.4
Theoretical system peak performance (GFlop/s)	22,938
Memory/node (GB)	8
Total memory (TB)	8.2

3

ASCI Q

LANL
LOS ALAMOS
HP ALPHASERVER SC
Rmax: 13.88 TFlops

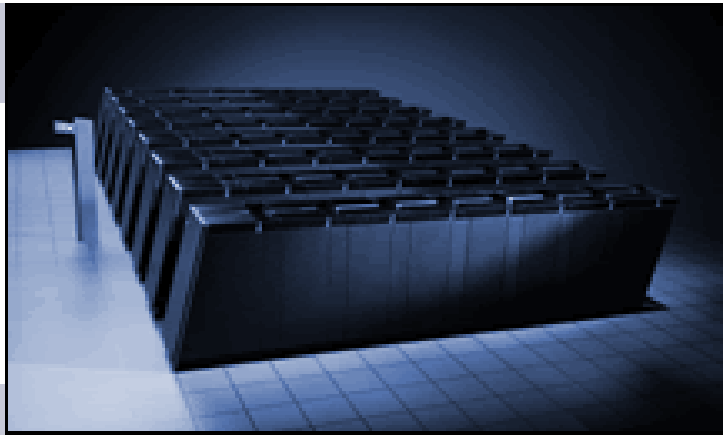
The Q supercomputing system at Los Alamos National Laboratory (LANL) is the most recent component of the Advanced Simulation and Computing (ASCI) program, a collaboration between the U. S. Department of Energy's National Nuclear Security Administration and the Sandia, Lawrence Livermore, and Los Alamos national laboratories. ASCI's mission is to create and use leading-edge capabilities in simulation and computational modeling. In an era without nuclear testing, these computational goals are vital for maintaining the safety

- 3072 AlphaServer ES45s from Hewlett Packard (formerly Compaq)
- 12,288 EV-68 1.25-GHz CPUs with 16-MB cache
- 33 Terabytes (TB) memory
- Gigabit fiber-channel disk drives providing 664 TB of global storage
- Dual controller accessible 72 GB drives arranged in 1536 5+1 RAID5 storage arrays, interconnected through fiber-channel switches to 384 file server nodes.

4

BlueGene

Thomas Watson Research Center
Yorktown Heights, USA
BlueGene/L DD1 Prototype
Rmax: 11.68 TFlops

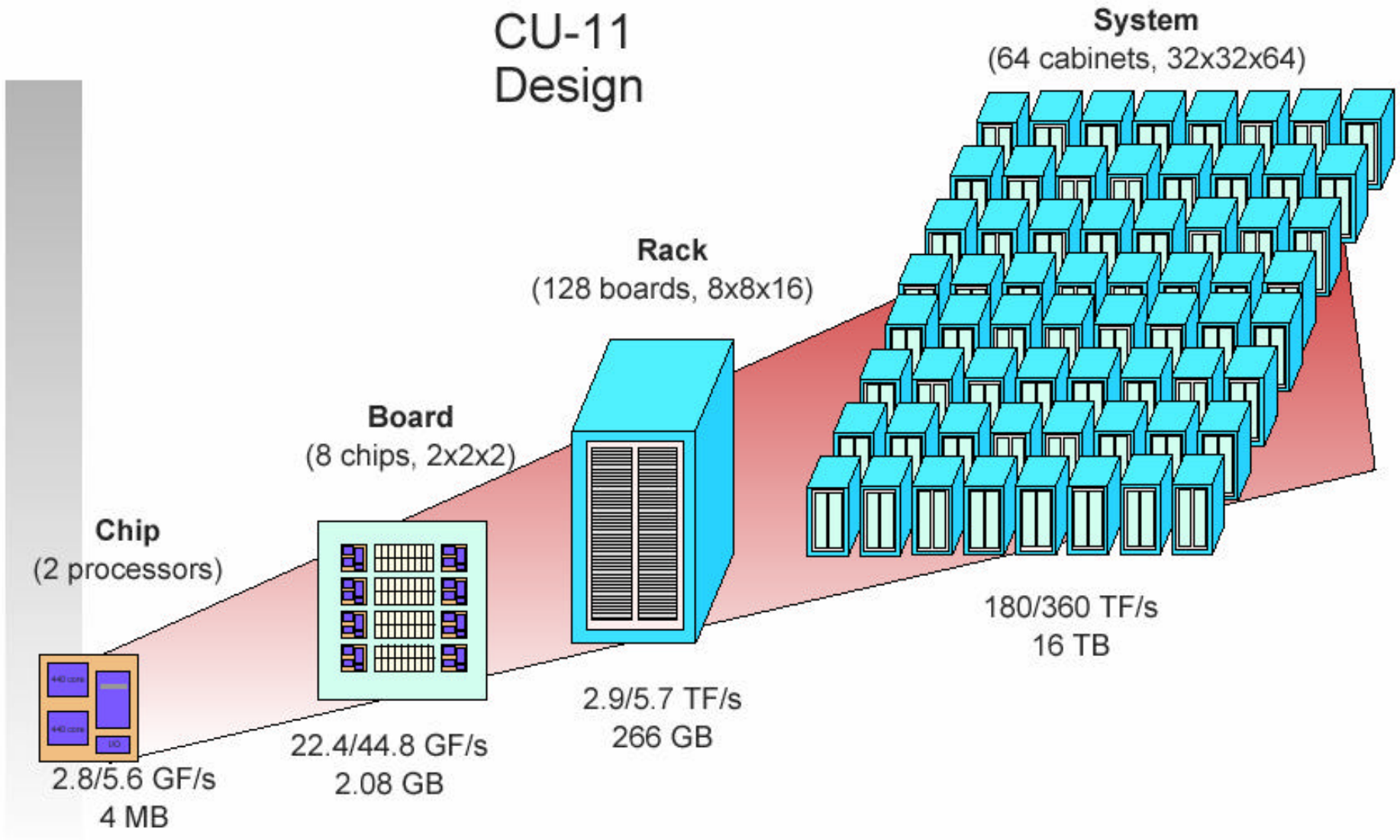


BlueGene/L is a cooperative project to design and build an innovative high-end computer architecture that can be scaled to the hundreds of teraflops (a trillion, or 1012 floating point operations per second). It is targeted for deployment in late 2004.

BlueGene/L is a scalable ultra-computer targeted for 216 = 65,536 compute nodes. BlueGene/L is a cellular architecture in that the basic building block of the system can be replicated in a regular pattern, with no introduction of bottlenecks as the system is scaled up.

Each BlueGene/L node consists of a single compute ASIC (an application-specific integrated circuit) and SDRAM-DDR memory chips. The compute ASIC is a complete system-on-a-chip, including all network interfaces and a modest amount of fast on-chip memory. An on-chip memory controller provides for access to larger external memory chips.

Blue Gene/L


CU-11
Design
System
(64 cabinets, 32x32x64)
Rack
(128 boards, 8x8x16)
Board
(8 chips, 2x2x2)
Chip
(2 processors)
180/360 TF/s
16 TB
2.9/5.7 TF/s
266 GB
22.4/44.8 GF/s
2.08 GB
2.8/5.6 GF/s
4 MB

5

Tungsten

NCSA
Urbana-Champaign, USA
Dell PowerEdge 1750/Myrinet
Rmax: 9.819 TFlops

Summary

Tungsten, NCSA's newest cluster, will employ more than 1,450 dual-processor Dell PowerEdge 1750 servers running Red Hat Linux, a Myrinet 2000 high-speed interconnect fabric, an I/O subcluster with more than 120 terabytes of DataDirect storage, and a dedicated 64-node applications development environment. Tungsten is expected to have a peak performance of 17.7 teraflops (17.7 trillion calculations per second).

Dell PowerEdge Servers - Microsoft Internet Explorer

Arquivo Editar Exibir Favoritos Ferramentas Ajuda

Endereço http://www1.us.dell.com/content/products/productdetails.aspx?c=us&cs=04&l=en&s=bsd&id=pedge_1750

Ir Links Assist

Buy Online or Call 1-800-917-DELL

DELL™ USA Small Business

Search

Advanced Search

Desktops | Notebooks | Servers, Storage & Networking | Printers | Software & Peripherals | Services & Support

My Account | My Cart | Order Status | Purchase Help

BACK TO: USA > Small Business > ... > Compare Rack Optimized Servers

E-mail page | Printable Version

PowerEdge 1750 Details

Questions? Dial and Ask. Call 1-800-551-3355 for assistance. More Details

PowerEdge 1750

~~From $1,499~~

Now from

$1,199

[Save 20% Instantly.]

Limited time offer.

Dell Business Credit

Convenient revolving credit. No minimum purchase. Apply and Buy!

Customize It

OR SELECT

Recommended Servers

- **Save up to 20% on select Systems & up to 25% on select Servers!**
  Big System & Server Savings - How Big is Up to You! Limited time offer. More Details
- **FREE Processor Upgrade!**
  With select PowerEdge™ servers. Expires 10/20/04. More Details
- **Understanding Dell™ Servers & Systems Management - only $195!**
  After $50 instant savings! Limited time offer. More Details

Tech Specs | Product Views | Product Highlights | Services

- Performance
- Expandability
- Additional Specs

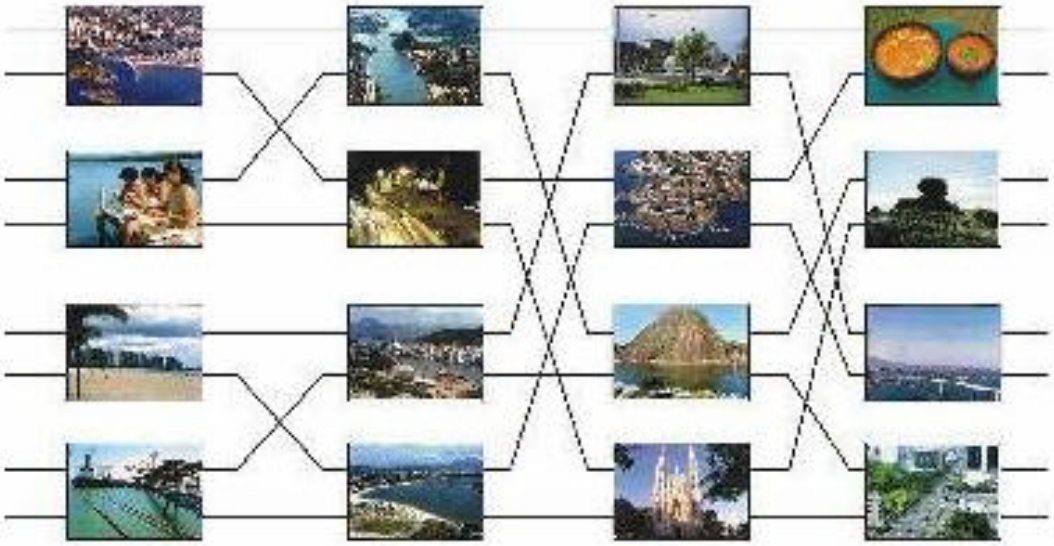
Processors

Dual Intel® Xeon® processors

- 3.2GHz, 533MHz front side bus, 2MB cache
- 3.2GHz, 533MHz front side bus, 1MB cache

Resources

- 360 Demo No Plug-in necessary
- PowerEdge 1750 Spec Sheet
- PowerEdge 1750 INFOBrief

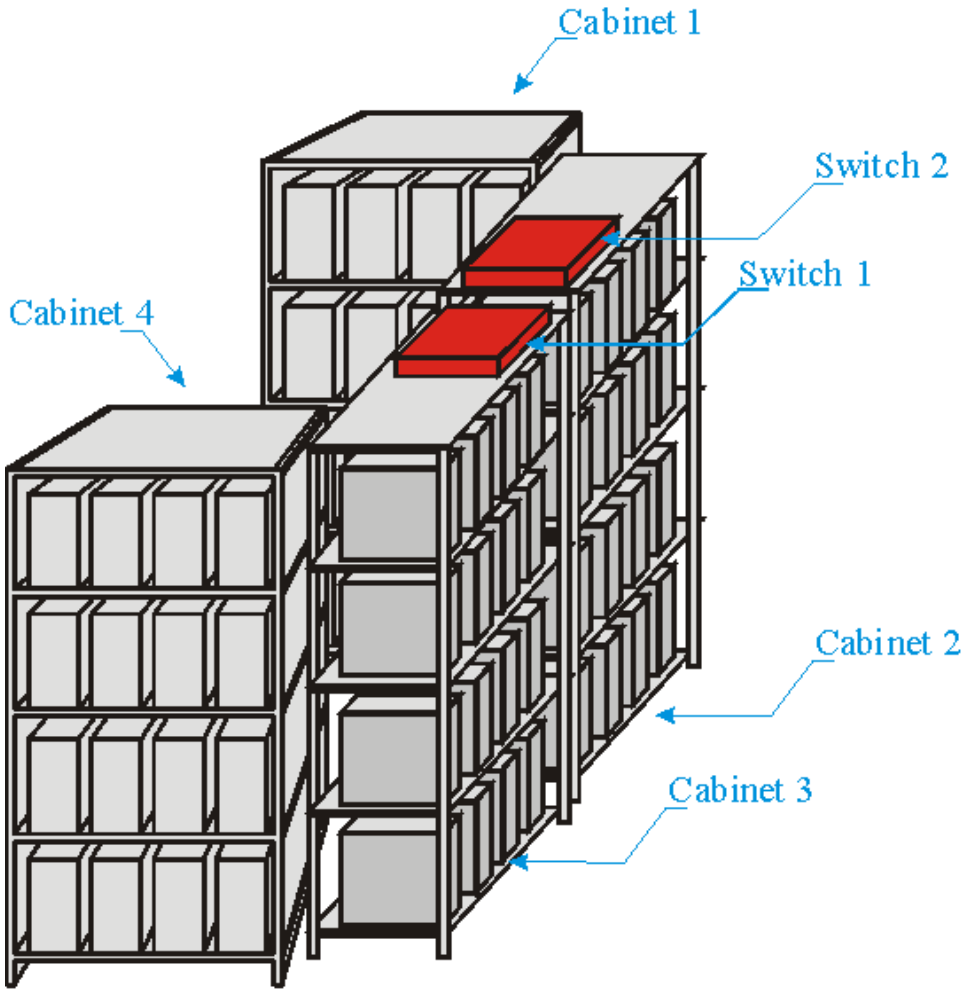


LCAD

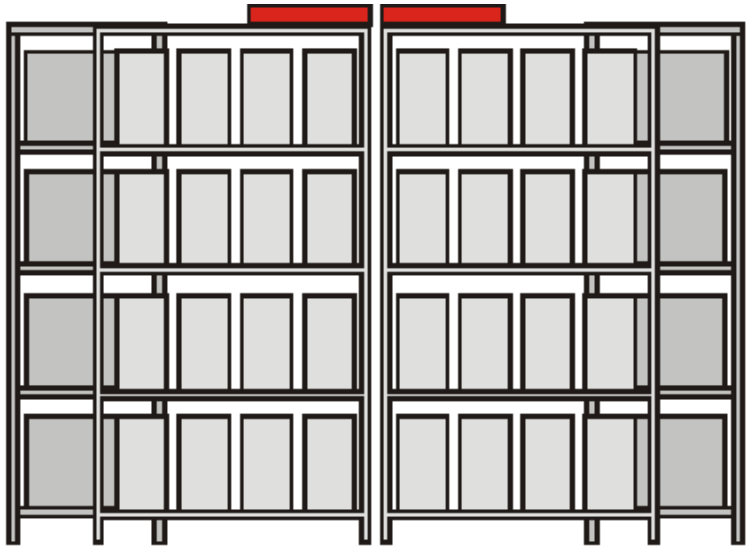
Laboratório de Computação de Alto Desempenho
Departamento de Informática
Centro Tecnológico
Universidade Federal do Espírito Santo


SAMSUNG

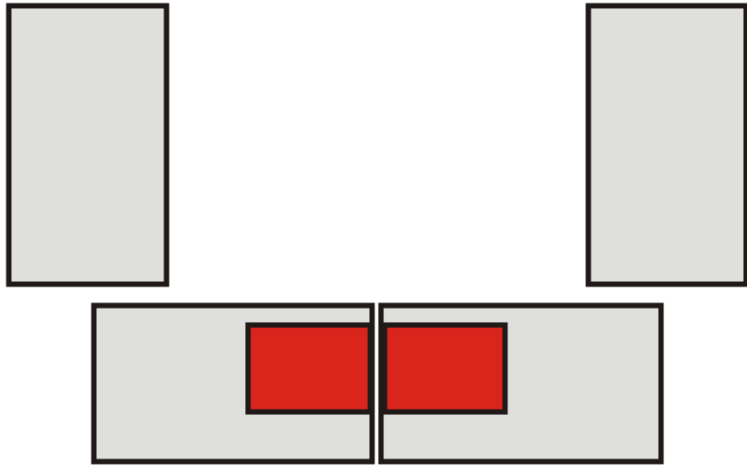

35
34
33
32
19
18
17
16
39
38
37
36
23
22
21
20
43
42
41
40
27
26
24



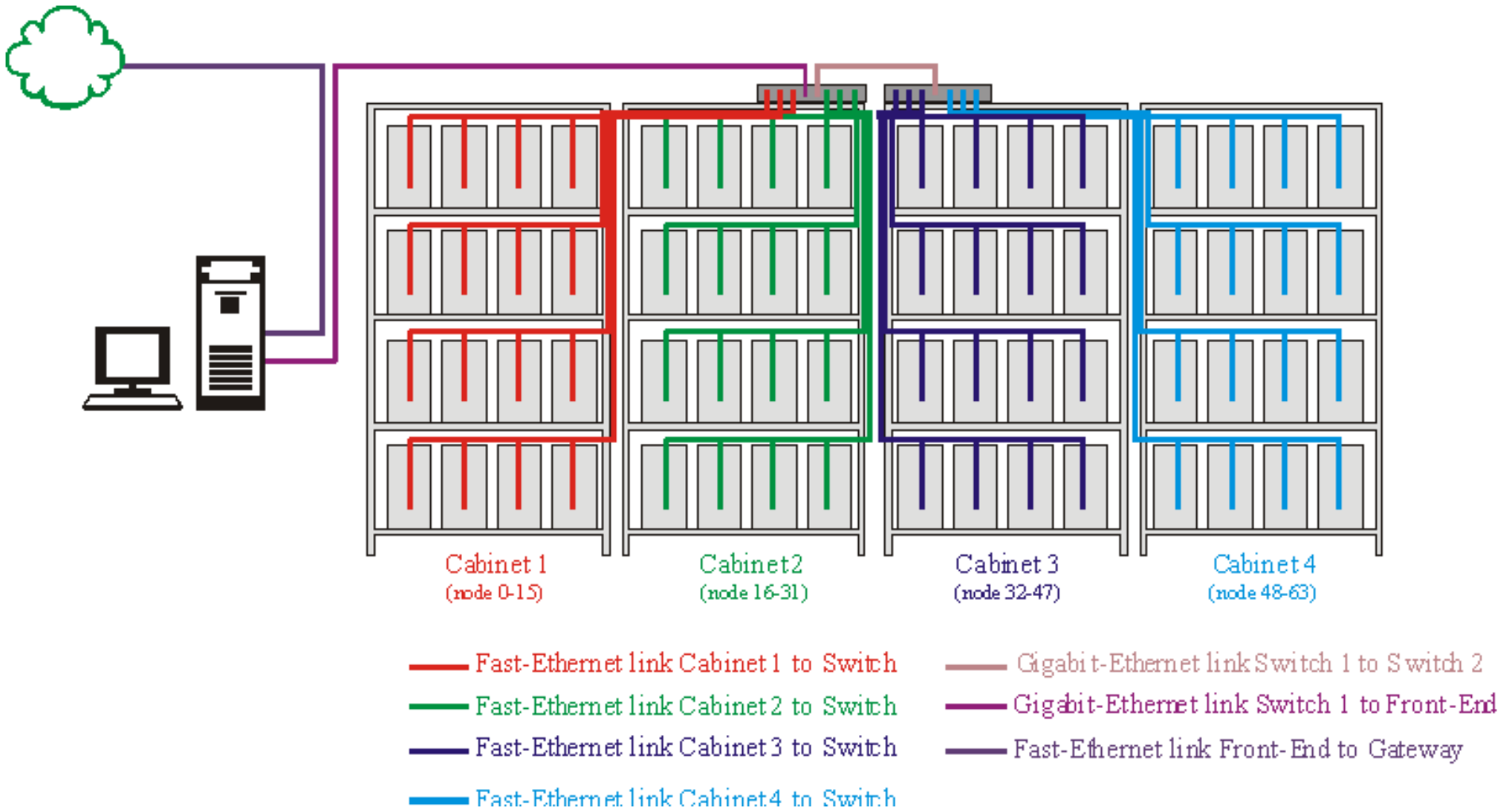
Perspective view



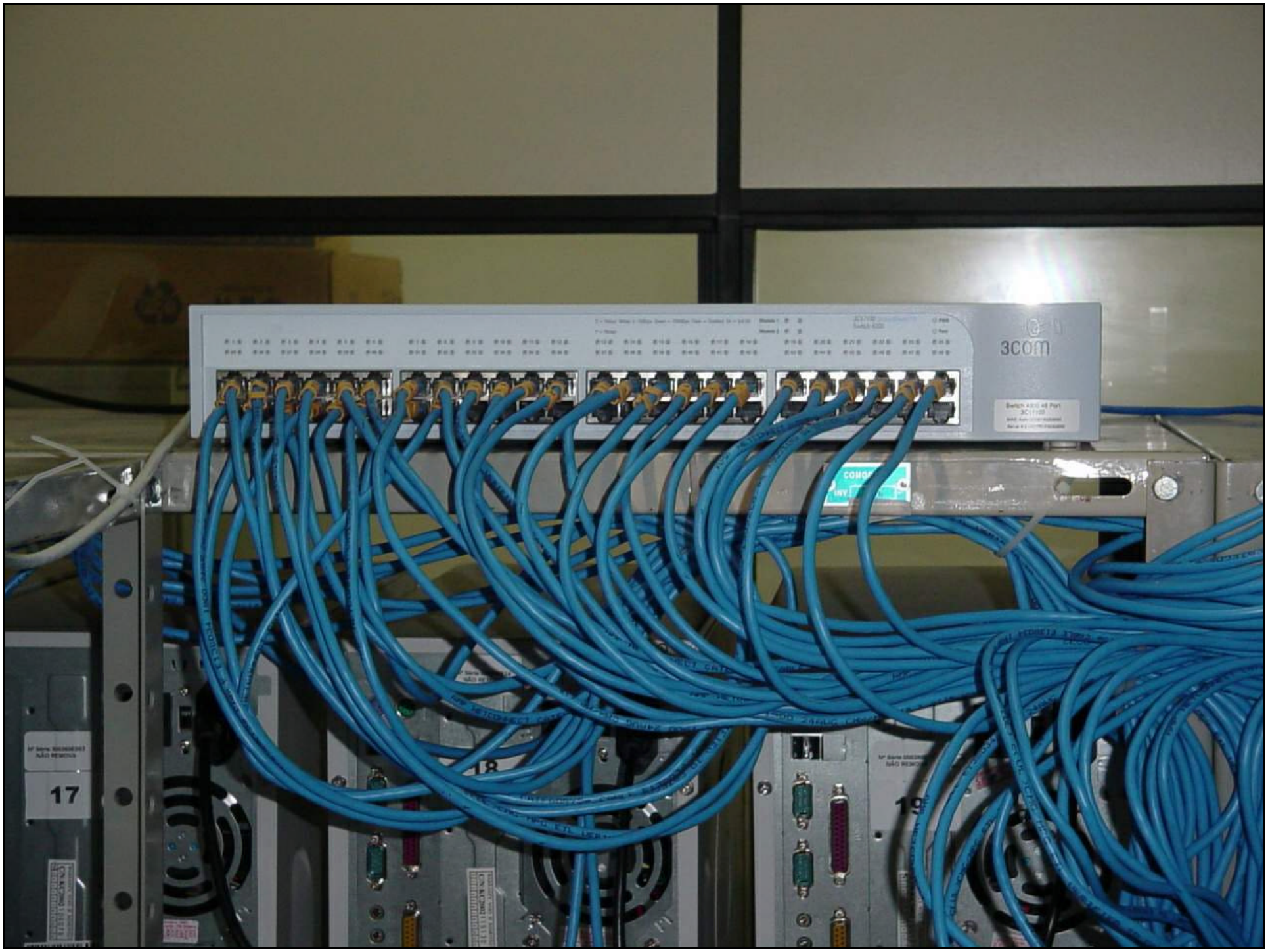
Front view



Top view


Cabinet 1
(node 0-15)
Cabinet2
(node 16-31)
Cabinet 3
(node 32-47)
Cabinet 4
(node 48-63)
Fast-Ethernet link Cabinet 1 to Switch
Fast-Ethernet link Cabinet 2 to Switch
Fast-Ethernet link Cabinet 3 to Switch
Fast-Ethernet link Cabinet4 to Switch
Gigabit-Ethernet link Switch 1 to Switch 2
Gigabit-Ethernet link Switch 1 to Front-End
Fast-Ethernet link Front-End to Gateway


34
51
38
54


3com
17

Totalmente operacional em janeiro de 2003, o Cluster do LCAD é composto de 64 nós de processamento e 1 nó de administração (master), com 256 MB de memória e 20Gb de capacidade de armazenamento por nó, o cluster totaliza 16 GB de memória RAM e 1,2 TB de capacidade de armazenamento.

Master: ATHLON XP 1800, 512 MB RAM, 90 GB de disco
2 placas de rede, uma 3COM Gigabit Ethernet para conexão com os nós de processamento e uma 3COM TX Fast-Ethernet para conexão externa
Linux Red Hat 7.1

Nós de processamento: 64 ATHLON's XP 1800, 256 RAM, 20 GB de disco, Placa de rede 3COM TX Fast-Ethernet Linux Red Hat 7.1 . Cada processador ATHLON XP 1800 possui 2 unidades de processamento para operações de ponto flutuando. Operando a 1,53 GHz tem-se uma performance teórica de pico de 3,06 Gflop/s por nó de processamento. Testes rodando o High-Performance Linpack Benchmark (programa para teste de desempenho de supercomputadores paralelos - basicamente uma fatoracao LU de uma matriz densa) mostraram uma performance de real de 1,79 Gflop/s por nó.

A performance teórica de pico do sistema é de 195,8 Gflop/s. Testes rodando o High-Performance Linpack Benchmark mostraram uma performance de real de 47,9 Gflop/s.

Consumo de Energia: Uma rede elétrica própria foi dimensionada especificamente para abastecer o cluster, o consumo total esta estimado em 100W por nó em funcionamento, isto é, aproximadamente 6800 W.

Refrigeração: O sistema está instalado em um ambiente de acusticamente isolado (o som dos ar-condicionados e de 65 ventiladores de CPU's é bastante intenso) - foto. A refrigeração é efetuada por dois ar-condicionados (totalizando 30.000 BTU's), mantendo redundância em caso de falha de um dos equipamentos. Adicionalmente, um sensor de temperatura conectado ao front-end monitora a temperatura do ambiente disparando um sinal de shutdown para os nós de processamento em caso de super-aquecimento. Devido a disposição em U do sistema existe a possibilidade de acumulo de calor no centro do U, assim um ventilador de teto foi posicionado diretamente sobre esta posição para garantir a circulação de ar no ambiente do cluster.

Configuração do sistema operacional e discos: O sistema operacional é instalado localmente em cada nó, uma vez que cada nó de processamento possui 20 Gb de disco rígido. A instalação local visa diminuir ao máximo o tráfego na rede. Somente os diretórios dos usuários estão no front-end (sistema NFS) para que os arquivos dos usuários estejam disponíveis em todos os nós.

Custo de montagem ======> **US$ 48.000,00**

Algoritmos Paralelos (disciplina regular)

Ementa:

1. Modelos de computação paralela:
   - Memória compartilhada: PRAM síncrona e assíncrona, QSP, QWQR, BSP, LogP, SMP
   - Memória distribuída: SMPD
   - circuitos combinatoriais
2. paradigmas de programação: multi-threaded, BSP, passagem de mensagem, funcional, tuple space, CSP
3. desempenho: profundidade de computação, trabalho, custo, speed-up, eficiência
4. Famílias fundamentais de algoritmos: árvore binária balanceada, divide-and-conquer, jumping pointer, compressão, randomização
5. aplicações
   - ordenação
   - grafos
   - processamento de strings
   - álgebra linear
   - otimização
   - Complexidade paralela: P-completeness

UFES
UNIVERSIDADE FEDERAL DO ESPÍRITO SANTO