CENTER FOR COMPUTATION & TECHNOLOGY
AT LOUISIANA STATE UNIVERSITY

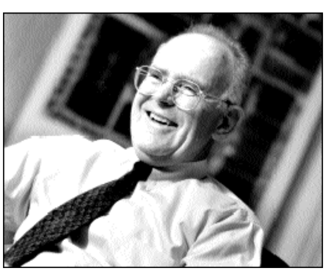
Cactus Retreat

An Overview of High Performance Computing, Clusters, and the Grid

Jack Dongarra
University of Tennessee
and
Oak Ridge National Laboratory

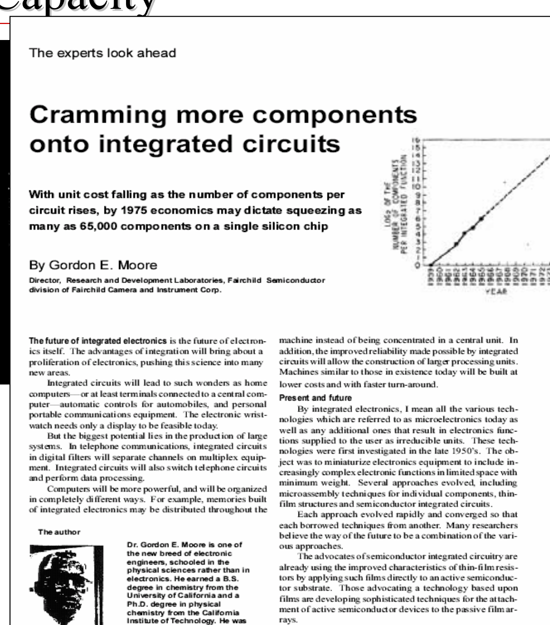
www.CactusCode.org

Technology Trends: Microprocessor Capacity



Gordon Moore
(co-founder of Intel)
Electronics Magazine, 1965
Number of devices/chip doubles every 18 months

2X transistors/Chip Every 1.5 years "Moore's Law"



The experts look ahead

Cramming more components onto integrated circuits

With unit cost falling as the number of components per circuit rises, by 1975 economics may dictate squeezing as many as 65,000 components on a single silicon chip

By Gordon E. Moore

Director, Research and Development Laboratories, Fairchild Semiconductor division of Fairchild Camera and Instrument Corp.

The future of integrated electronics is the future of electronics itself. The advantages of integration will bring about a proliferation of electronics, pushing this science into many new areas.

Integrated circuits will lead to such wonders as home computers—or at least terminals connected to a central computer—automatic controls for automobiles, and personal portable communications equipment. The electronic wristwatch needs only a display to be feasible today.

But the biggest potential lies in the production of large systems. In telephone communications, integrated circuits in digital filters will separate channels on multiplex equipment. Integrated circuits will also switch telephone circuits and perform data processing.

Computers will be more powerful, and will be organized in completely different ways. For example, memories built of integrated electronics may be distributed throughout the machine instead of being concentrated in a central unit. In addition, the improved reliability made possible by integrated circuits will allow the construction of larger processing units. Machines similar to those in existence today will be built at lower costs and with faster turn-around.

Present and future

By integrated electronics, I mean all the various technologies which are referred to as microelectronics today as well as any additional ones that result in electronics functions supplied to the user as irreducible units. These technologies were first investigated in the late 1950's. The object was to miniaturize electronics equipment to include increasingly complex electronic functions in limited space with minimum weight. Several approaches evolved, including microassembly techniques for individual components, thin-film structures and semiconductor integrated circuits.

Each approach evolved rapidly and converged so that each borrowed techniques from another. Many researchers believe the way of the future to be a combination of the various approaches.

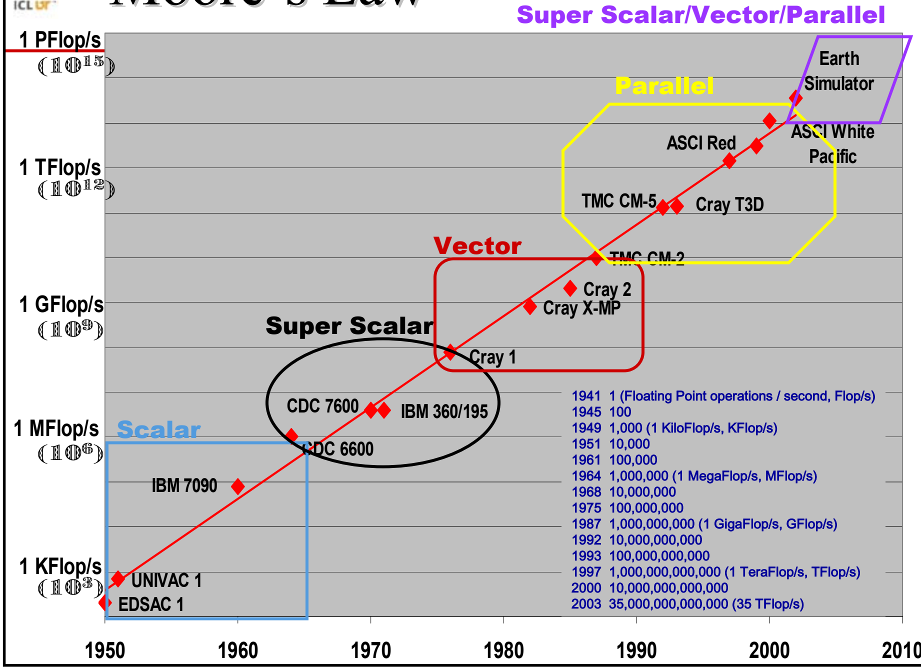
The advocates of semiconductor integrated circuitry are already using the improved characteristics of thin-film resistors by applying such films directly to an active semiconductor substrate. Those advocating a technology based upon films are developing sophisticated techniques for the attachment of active semiconductor devices to the passive film arrays.

Both approaches have worked well and are being used

The author

Dr. Gordon E. Moore is one of the new breed of electronic engineers, schooled in the physical sciences rather than in electronics. He earned a B.S. degree in chemistry from the University of California and a Ph.D. degree in physical chemistry from the California Institute of Technology. He was one of the founders of Fairchild

Moore's Law

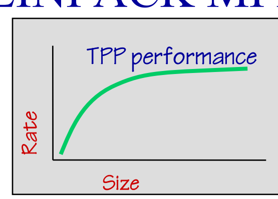


TOP500 superCOMPUTER

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

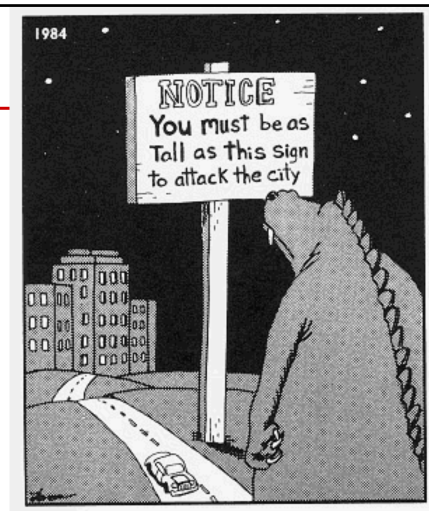
  SC'xy in the States in November

  Meeting in Mannheim, Germany in June

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

What is a Supercomputer?

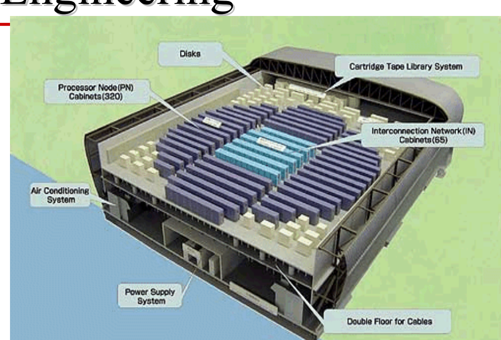
- A supercomputer is a hardware and software system that provides close to the maximum performance that can currently be achieved.
- Over the last 10 years the range for the Top500 has increased greater than Moore's Law
- 1993:
  - #1 = 59.7 GFlop/s
  - #500 = 422 MFlop/s
- 2003:
  - #1 = 35.8 TFlop/s
  - #500 = 403 GFlop/s



Why do we need them? Computational fluid dynamics, protein folding, climate modeling, national security, in particular for cryptanalysis and for simulating nuclear weapons to name a few.

A Tour de Force in Engineering

- Homogeneous, Centralized, Proprietary, Expensive!
- Target Application: CFD-Weather, Climate, Earthquakes
- 640 NEC SX/6 Nodes (mod)
  - 5120 CPUs which have vector ops
  - Each CPU 8 Gflop/s Peak
- 40 TFlop/s (peak)
- ~ 1/2 Billion $ for machine, software, & building
- Footprint of 4 tennis courts
- 7 MWatts
  - Say 10 cent/KWhr - $16.8K/day = $6M/year!
- Expect to be on top of Top500 until 60-100 TFlop ASCI machine arrives
- From the Top500 (November 2003)
  - Performance of ESC
    - > Σ Next Top 3 Computers

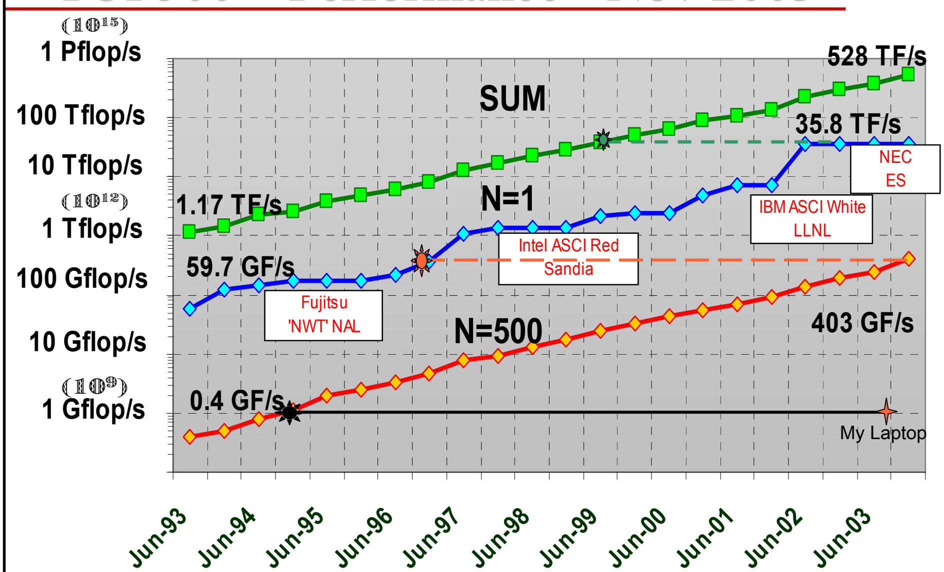


November 2003

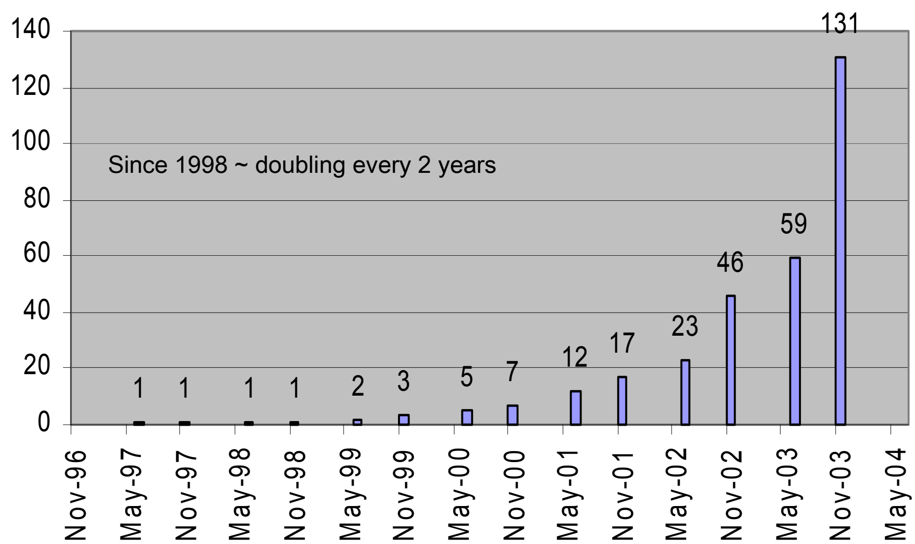
	Manufacturer	Computer	Rmax Tflop/s	Installation Site	Year	# Proc	Rpeak Tflop/s
1	NEC	Earth-Simulator	35.8	Earth Simulator Center Yokohama	2002	5120	40.90
2	Hewlett-Packard	ASCI Q - AlphaServer SC ES45/1.25 GHz	13.9	Los Alamos National Laboratory Los Alamos	2002	8192	20.48
3	Self	Apple G5 Power PC w/Infiniband 4X	10.3	Virginia Tech Blacksburg, VA	2003	2200	17.60
4	Dell	PowerEdge 1750 P4 Xeon 3.6 Ghz w/Myrinet	9.82	University of Illinois U/C Urbana/Champaign	2003	2500	15.30
5	Hewlett-Packard	rx2600 Itanium2 1 GHz Cluster – w/Quadrics	8.63	Pacific Northwest National Laboratory Richland	2003	1936	11.62
6	Linux NetworX	Opteron 2 GHz, w/Myrinet	8.05	Lawrence Livermore National Laboratory Livermore	2003	2816	11.26
7	Linux NetworX	MCR Linux Cluster Xeon 2.4 GHz – w/Quadrics	7.63	Lawrence Livermore National Laboratory Livermore	2002	2304	11.06
8	IBM	ASCI White, Sp Power3 375 MHz	7.30	Lawrence Livermore National Laboratory Livermore	2000	8192	12.29
9	IBM	SP Power3 375 MHz 16 way	7.30	NERSC/LBNL Berkeley	2002	6656	9.984
10	IBM	xSeries Cluster Xeon 2.4 GHz – w/Quadrics	6.59	Lawrence Livermore National Laboratory Livermore	2003	1920	9.216

50% of top500 performance in top 9 machines; 131 system > 1 TFlop/s; 210 machines are clusters

TOP500 – Performance - Nov 2003



Number of Systems on Top500 > 1 Tflop/s Over Time



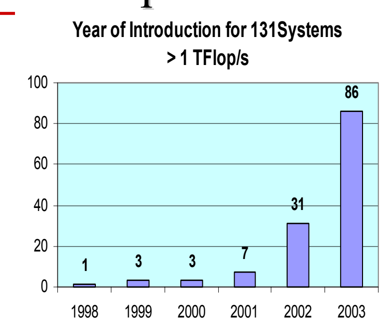
Factoids on Machines > 1 TFlop/s

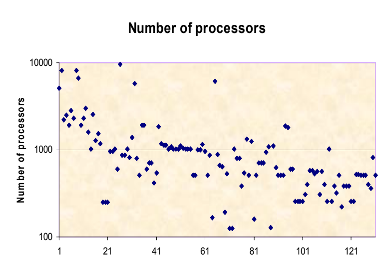
- 131 Systems
- 80 Clusters (61%)

- Average rate: 2.44 Tflop/s
- Median rate: 1.55 Tflop/s

- Sum of processors in Top131: 155,161
  - Sum for Top500: 267,789
- Average processor count: 1184
- Median processor count: 706

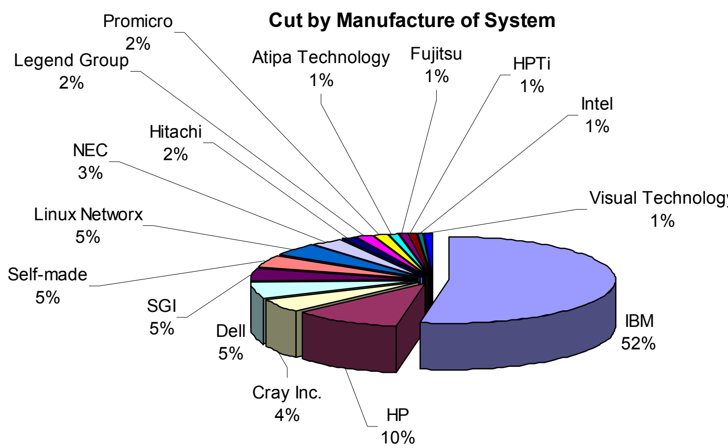
- Numbers of processors
  - Most number of processors: 9632_{26}
    - ASCI Red
  - Fewest number of processors: 124_{71}
    - Cray X1





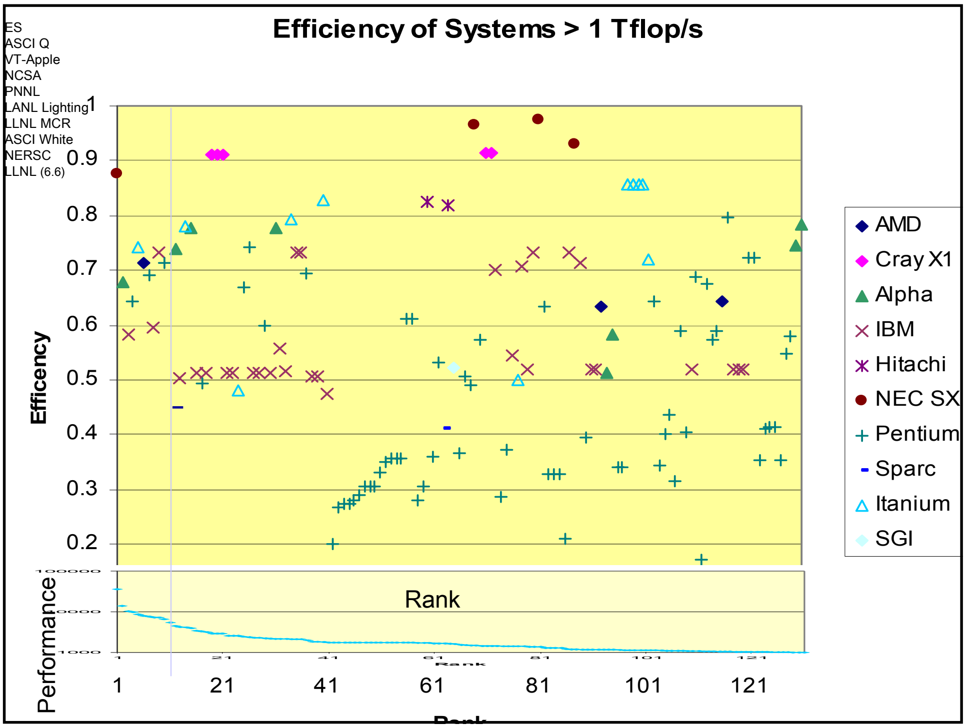
Percent Of 131 Systems Which Use The Following Processors > 1 TFlop/s

About a half are based on 32 bit architecture
9 (11) Machines have a Vector instruction Sets



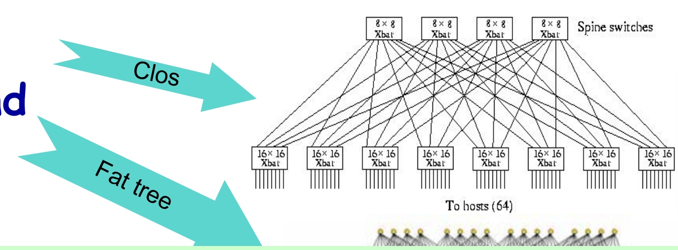
What About Efficiency?

- Talking about Linpack
- What should be the efficiency of a machine on the Top131 be?
  - Percent of peak for Linpack
    - 90% ?
    - 80% ?
    - 70% ?
    - 60% ?
  - …
- Remember this is $O(n^3)$ ops and $O(n^2)$ data
  - Mostly matrix multiply

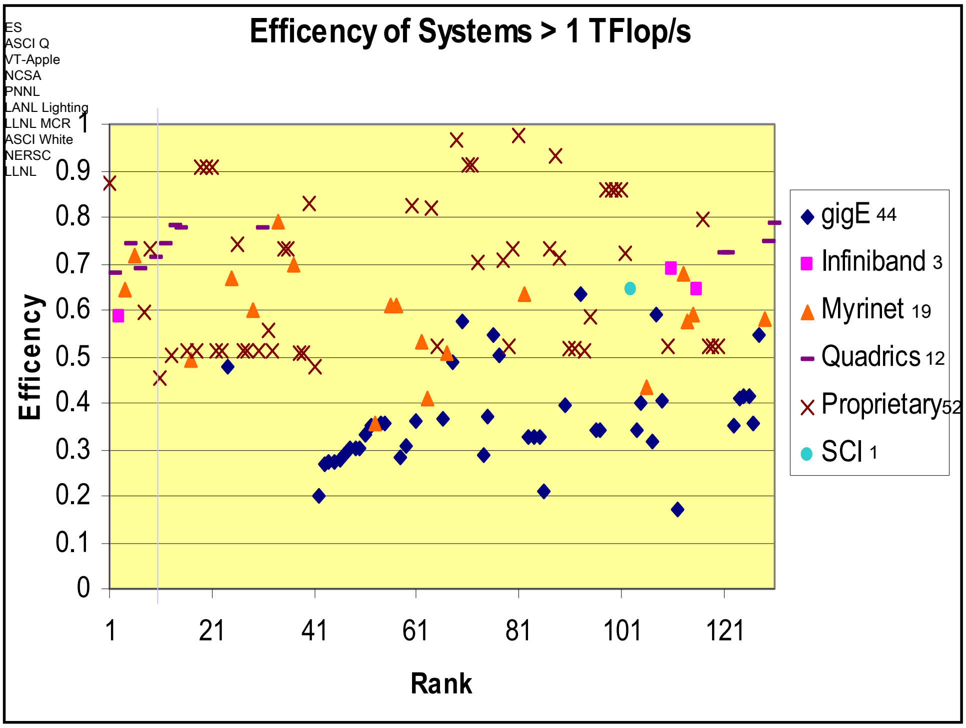


Commodity Interconnects

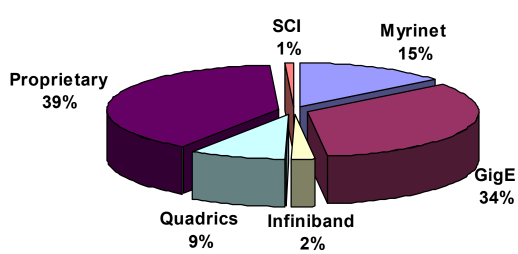
- Gig Ethernet
- Myrinet
- Infiniband
- QsNet
- SCI



	Switch topology	$ NIC	$Sw/node	$ Node	Lt(us)/BW (MB/s) (MPI)
Gigabit Ethernet	Bus	$ 50	$ 50	$ 100	30 / 100
SCI	Torus	$1,600	$ 0	$1,600	5 / 300
QsNetII	Fat Tree	$1,200	$1,700	$2,900	3 / 880
Myrinet (D card)	Clos	$ 700	$ 400	$1,100	6.5/ 240
IB 4x	Fat Tree	$1,000	$ 400	$1,400	6 / 820



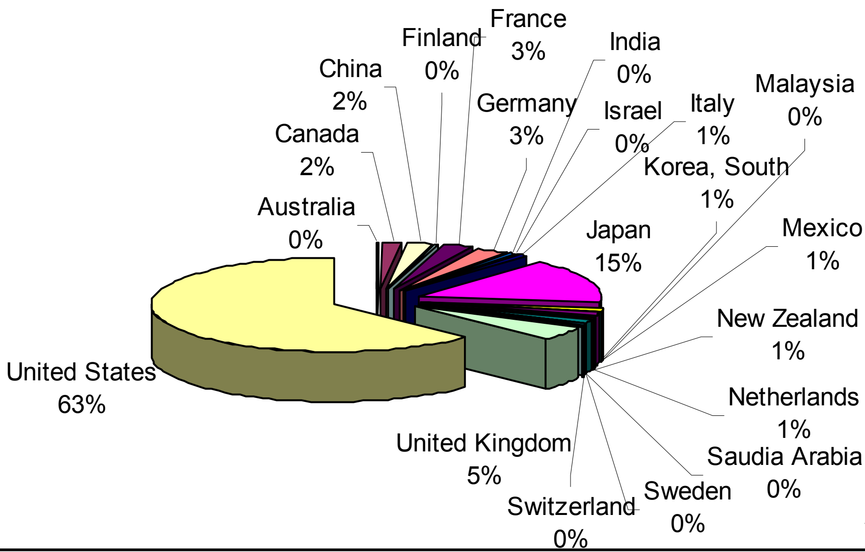
Interconnects Used



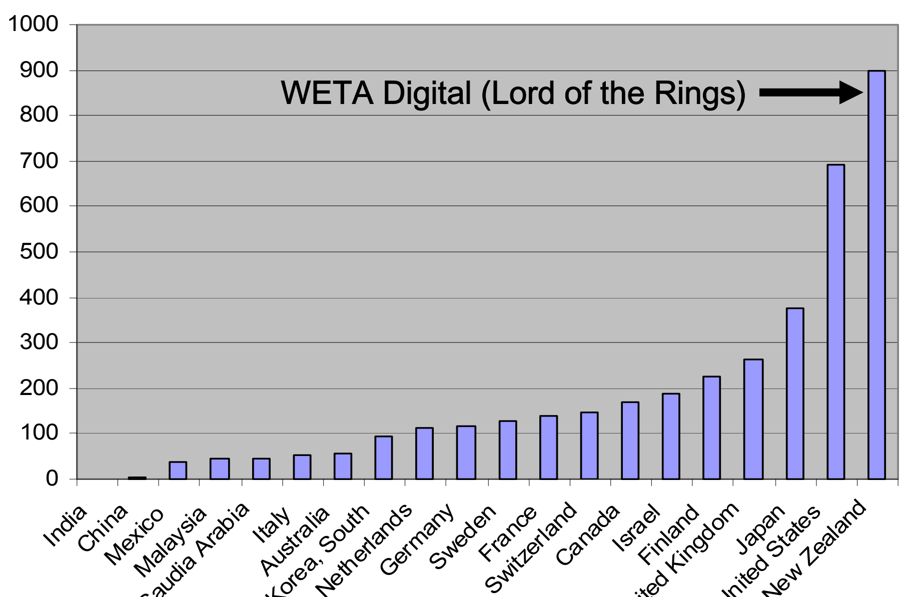
Efficiency for Linpack

	Largest node count	min	max	average
GigE	1024	17%	63%	37%
SCI	120	64%	64%	64%
QsNetII	2000	68%	78%	74%
Myrinet	1250	36%	79%	59%
Infiniband 4x	1100	58%	69%	64%
Proprietary	9632	45%	98%	68%

Country Percent by Total Performance

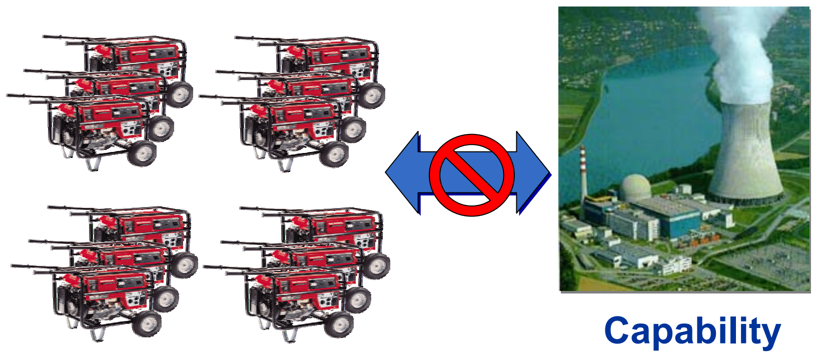


KFlop/s per Capita (Flops/Pop)



A Tool and A Market for Every Task

200K Honda units at 5 KW to equal a 1 GW nuclear plant

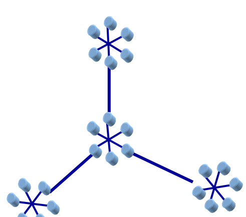


Capability

- **Each targets different applications**
  - understand application needs

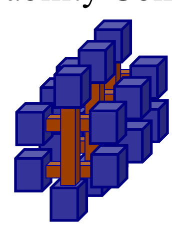
Taxonomy

Cluster Computing



- Commodity processors and switch
- Processors design point for web servers & home pc's
- Leverage millions of processors
- Price point appears attractive for scientific computing

Capability Computing



- Special purpose processors and interconnect
- High Bandwidth, low latency communication
- Designed for scientific computing
- Relatively few machines will be sold
- High price

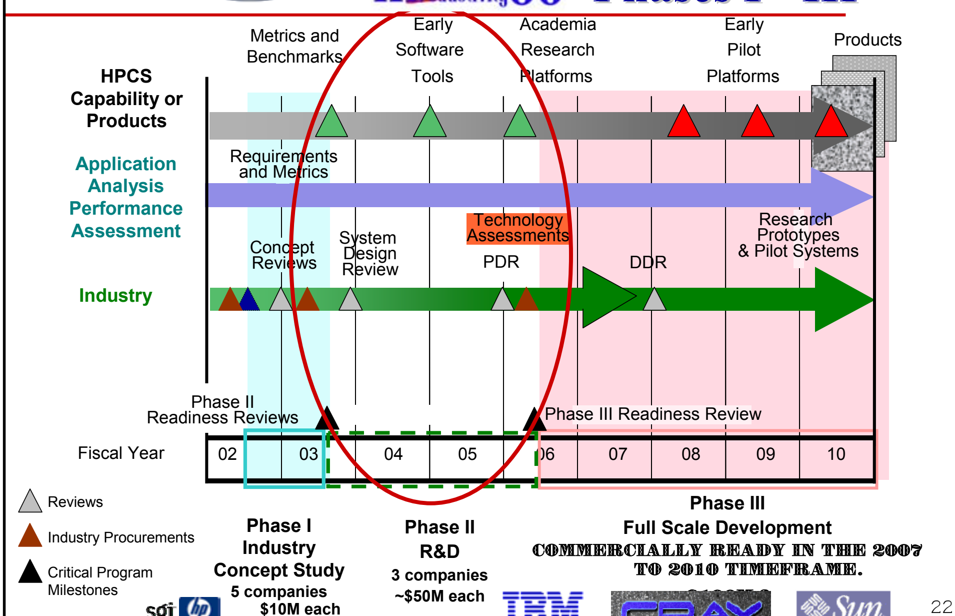
High Bandwidth vs Commodity Systems

- High bandwidth systems have traditionally been vector computers
  - Designed for scientific problems
  - Capability computing
- Commodity processors are designed for web servers and the home PC market

  (should be thankful that the manufactures keep the 64 bit fl pt)
  - Used for cluster based computers leveraging price point
- Scientific computing needs are different
  - Require a better balance between data movement and floating point operations. Results in greater efficiency.

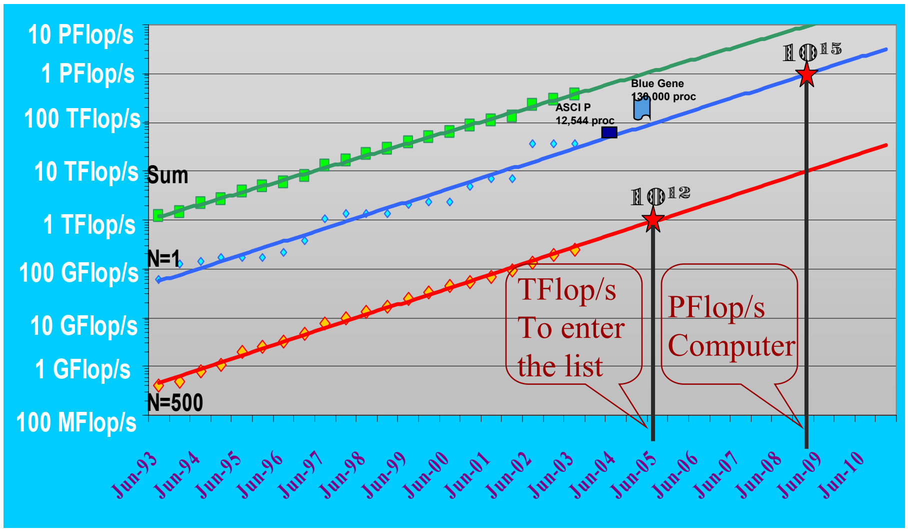
	Earth Simulator (NEC)	Cray X1 (Cray)	ASCI Q (HP EV68)	MCR (Dual Xeon)	VT Big Mac (Dual IBM PPC)
Year of Introduction	2002	2003	2002	2002	2003
Node Architecture	Vector	Vector	Alpha	Pentium	Power PC
Processor Cycle Time	500 MHz	800 MHz	1.25 GHz	2.4 GHz	2 GHz
Peak Speed per Processor	8 Gflop/s	12.8 Gflop/s	2.5 Gflop/s	4.8 Gflop/s	8 Gflop/s
Bytes/flop (main memory)	4	2.6	0.8	0.44	0.5

HPCS Phases I - III

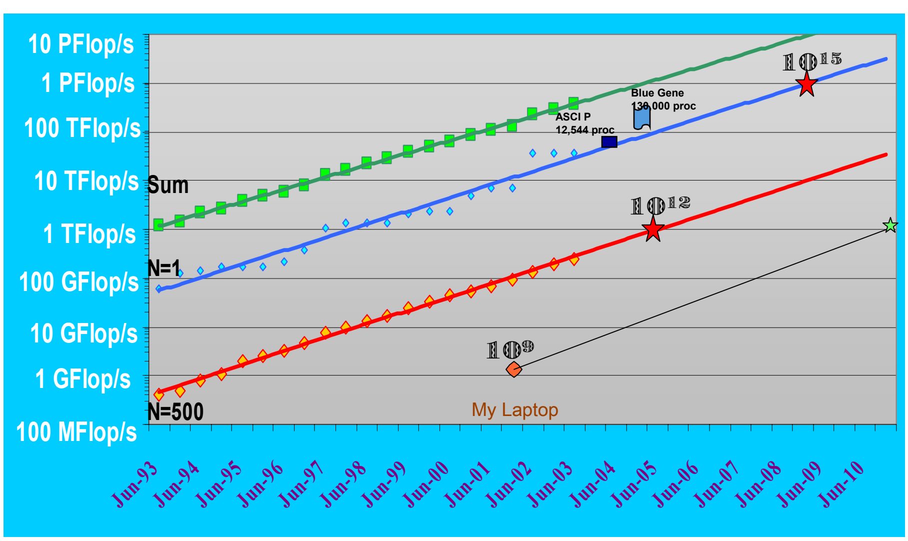


22

Performance Extrapolation

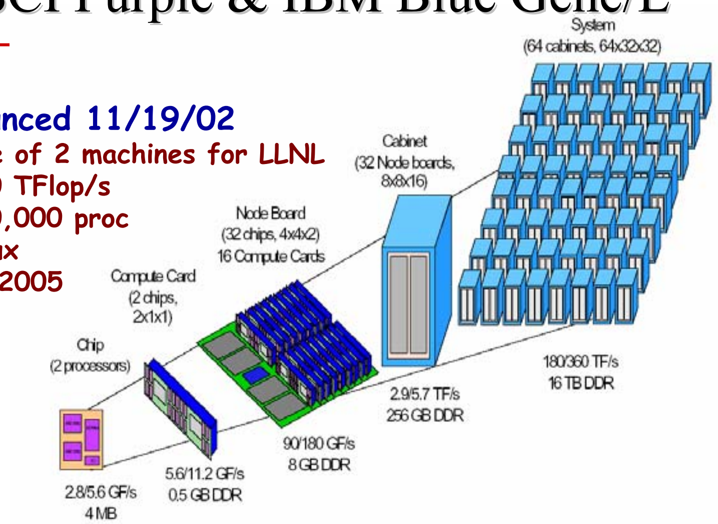


Performance Extrapolation



ASCI Purple & IBM Blue Gene/L

- Announced 11/19/02
  - One of 2 machines for LLNL
  - 360 TFlop/s
  - 130,000 proc
  - Linux
  - FY 2005

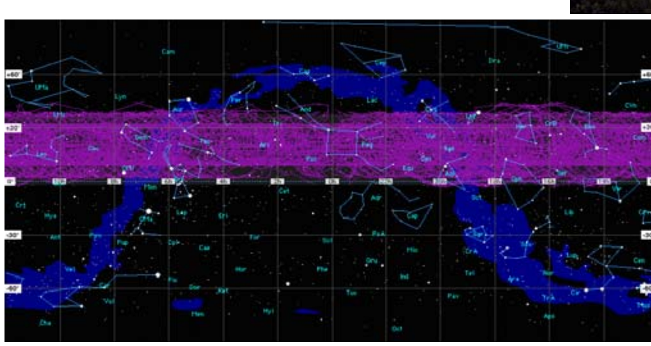


  - Preliminary machine
    - IBM Research BlueGene/L
      - PowerPC 440, 500MHz w/custom proc/interconnect
      - 512 Nodes (1024 processors)
      - 1.435 Tflop/s (2.05 Tflop/s Peak)

Plus
ASCI Purple
IBM Power 5 based
12K proc, 100 TFlop/s

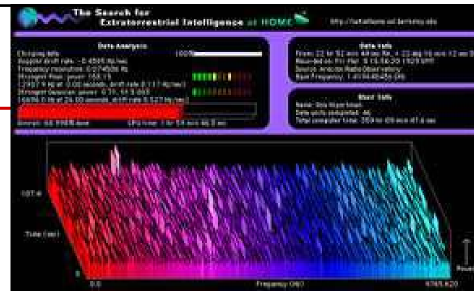
SETI@home: Global Distributed Computing

- Running on 500,000 PCs, ~1300 CPU Years per Day
  - 1.3M CPU Years so far
- Sophisticated Data & Signal Processing Analysis
- Distributes Datasets from Arecibo Radio Telescope

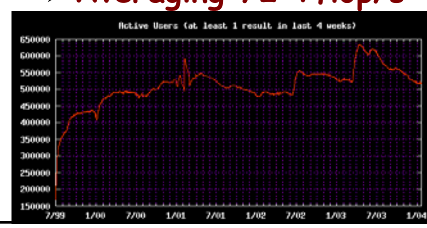


SETI@home

- Use thousands of Internet-connected PCs to help in the search for extraterrestrial intelligence.
- When their computer is idle or being wasted this software will download ~ half a MB chunk of data for analysis. Performs about 3 Tflops for each client in 15 hours.
- The results of this analysis are sent back to the SETI team, combined with thousands of other participants.
- About 5M users

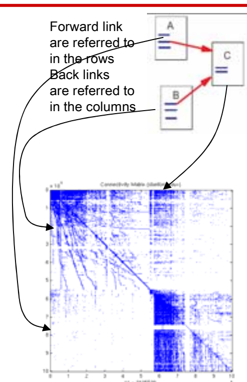


- Largest distributed computation project in existence
  - Averaging 72 Tflop/s



Google

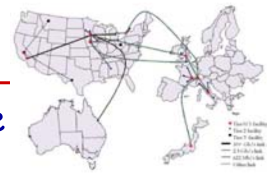
- Google query attributes
  - 150M queries/day (2000/second)
  - 100 countries
  - 3.3B documents in the index
- Data centers
  - 100,000 Linux systems in data centers around the world
    - 15 TFlop/s and 1000 TB total capability
    - 40-80 1U/2U servers/cabinet
    - 100 MB Ethernet switches/cabinet with gigabit Ethernet uplink
  - growth from 4,000 systems (June 2000)
    - 18M queries then
- Performance and operation
  - simple reissue of failed commands to new servers
  - no performance debugging
    - problems are not reproducible

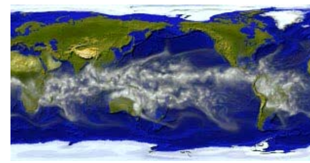


Eigenvalue problem
$n=3.3 \times 10^9$
(see: MathWorks Cleve's Corner)

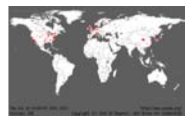
Source: Monika Henzinger, Google & Cleve Moler

Science and Technology





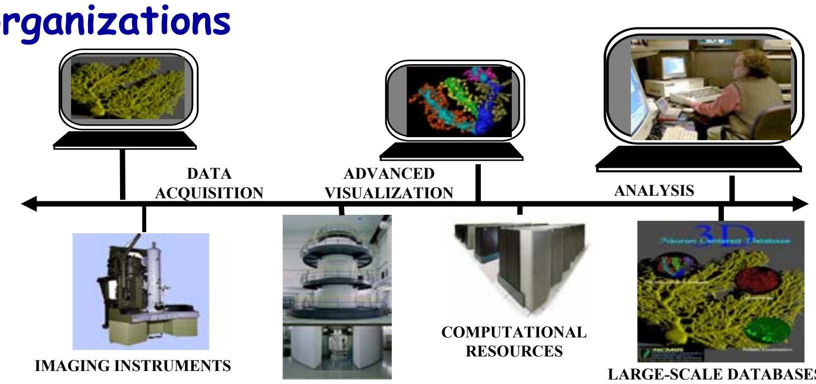
- Today, large science projects are conducted by global teams using sophisticated combinations of
  - *Computers*
  - *Networks*
  - *Visualization*
  - *Data storage*
  - *Remote instruments*
  - *People*
  - *Other resources*



- ***Information Infrastructure provides a way to integrate resources to support modern applications***

Grid Computing is About …

Resource sharing & coordinated problem solving in dynamic, multi-institutional virtual organizations



The most pressing scientific challenges require application solutions that are multidisciplinary and multi-scale.

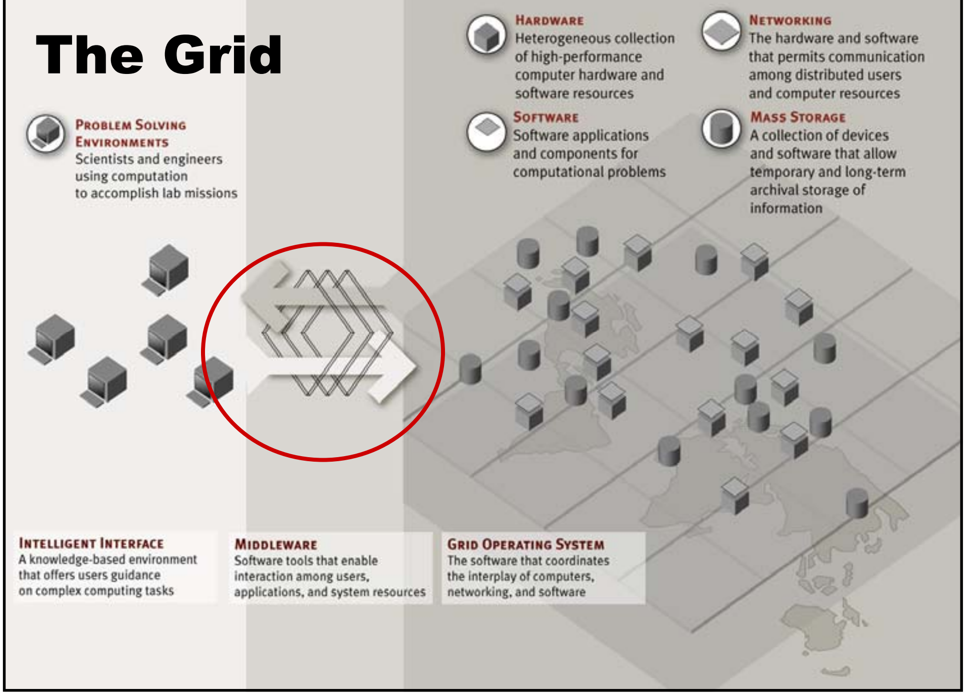
The Grid

- Motivation: When communication is close to free we should not be restricted to local resources when solving problems.
- Infrastructure that builds on the Internet and the Web
- Enable and exploit large scale sharing of resources
- Virtual organization
  - Loosely coordinated groups
- Provides for remote access of resources
  - Scalable
  - Secure
  - Reliable mechanisms for discovery and access

Grid Software Challenges

- Simplified programming
  - reduced complexity and coordination
- Accounting and resource economies
  - "non-traditional" resources and concurrency
    - shared resource costs and denial of service
  - negotiation and equilibration
    - exchange rates and sharing
- Scheduling and adaptation
  - performance, fault-tolerance, and access
    - networks, computing, storage, and sensors
- On-demand access
  - unique observational events and sensor fusion
    - "instant" access and nimble scheduling
- Managing bandwidth and latency
  - lambda dominance and exploitation

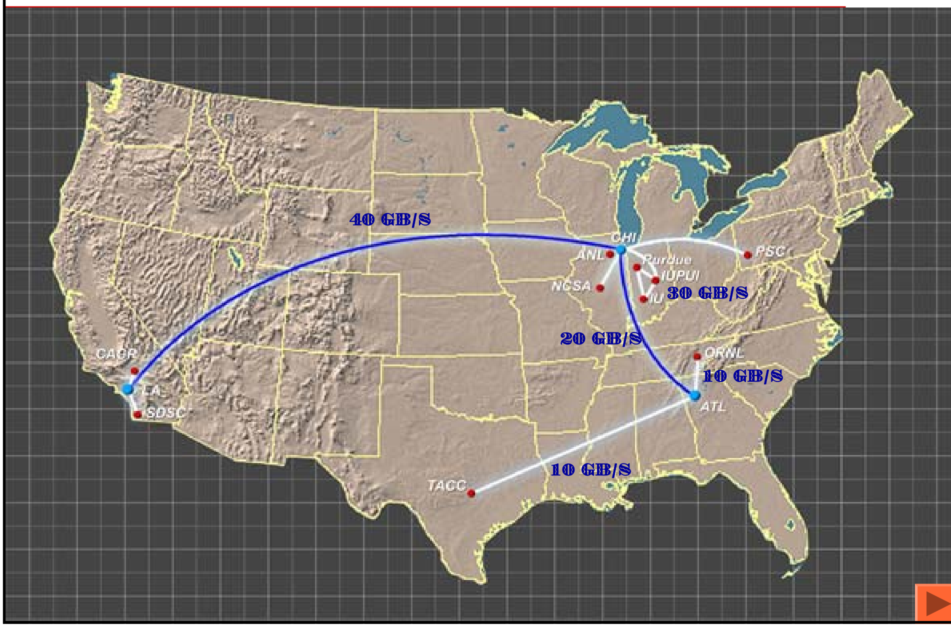
The Grid



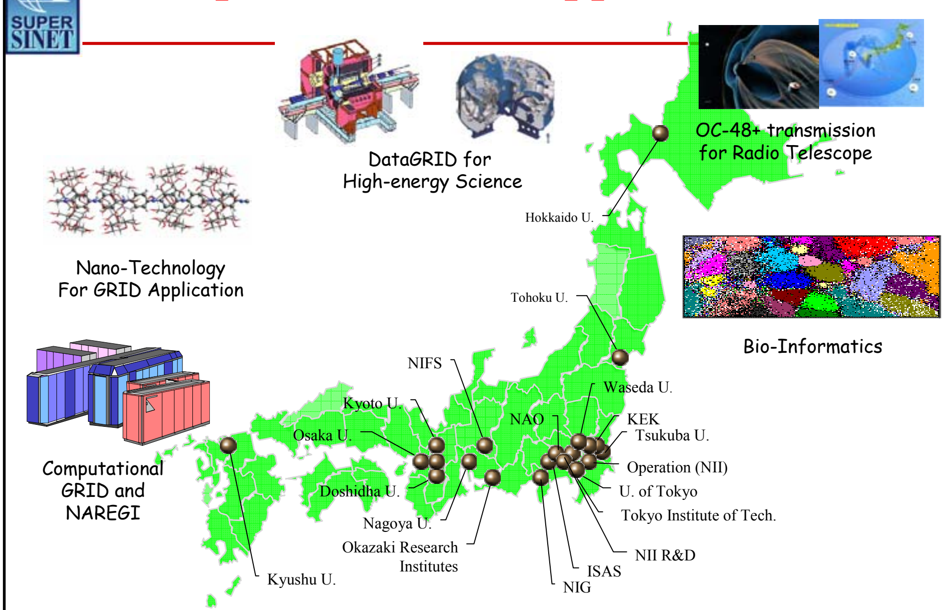
Science Grid Projects

TeraGrid 2003

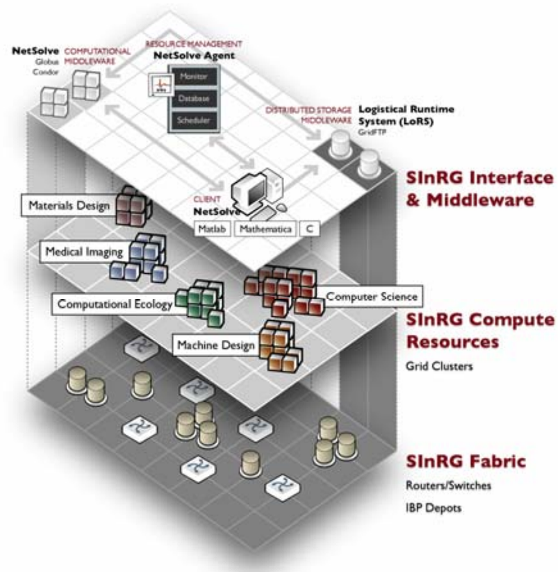
Prototype for a National Cyberinfrastructure

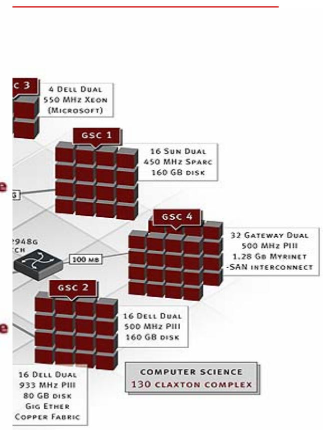


SuperSINET and Applications



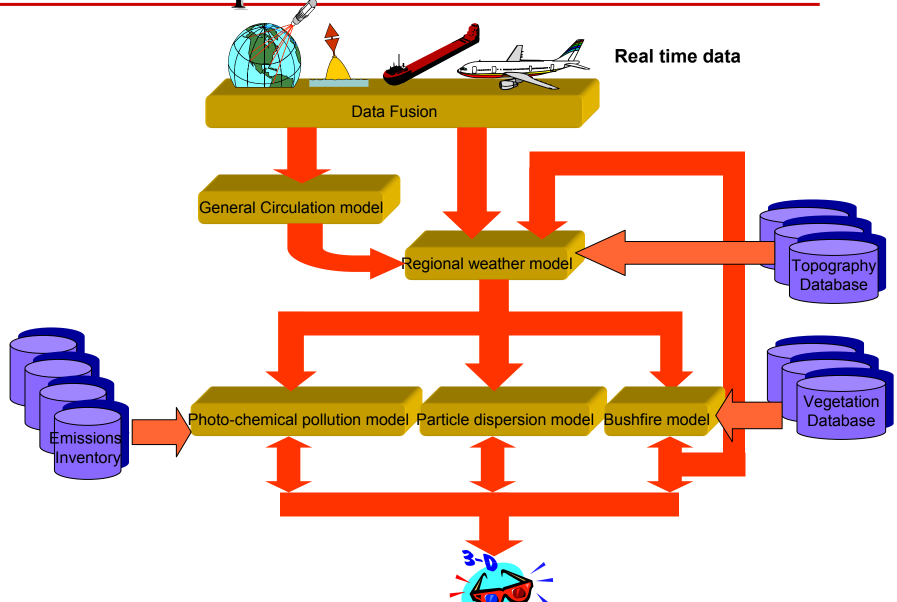
University of Tennessee Deployment: Scalable Intracampus Research Grid: SInRG



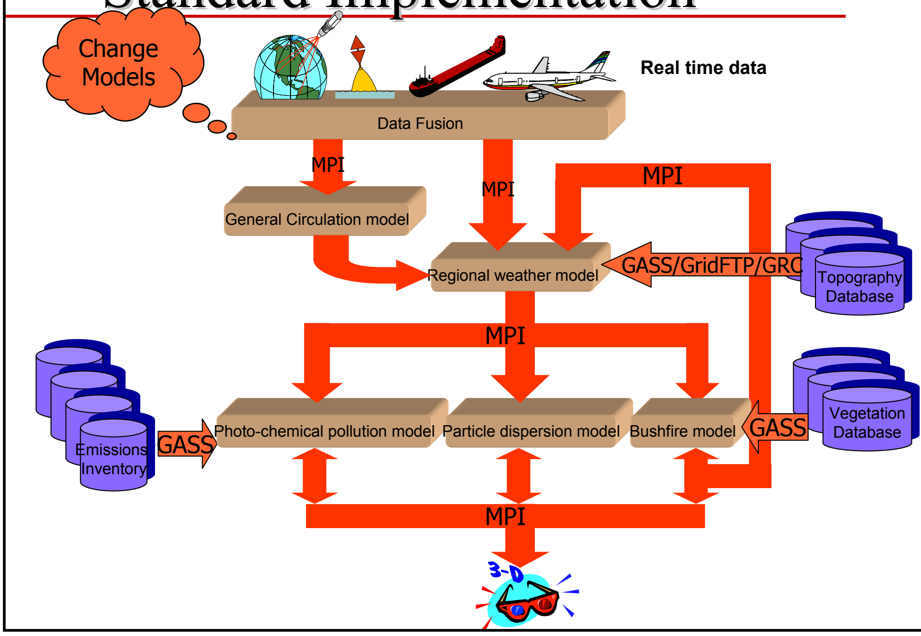


- Federated Ownership: CS, Chem Eng., Medical School, Computational Ecology, El. Eng.
- Real applications, middleware development, logistical networking

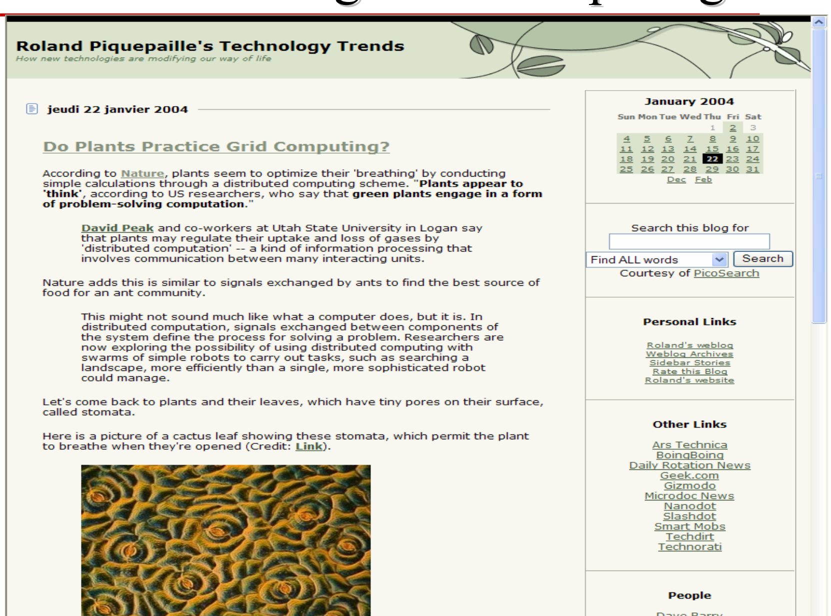
Atmospheric Sciences Grid



Standard Implementation



Are Plants Doing Grid Computing?



Roland Piquepaille's Technology Trends

How new technologies are modifying our way of life

jeudi 22 janvier 2004

Do Plants Practice Grid Computing?

According to Nature, plants seem to optimize their 'breathing' by conducting simple calculations through a distributed computing scheme. "**Plants appear to 'think'**, according to US researchers, who say that **green plants engage in a form of problem-solving computation.**"

> David Peak and co-workers at Utah State University in Logan say that plants may regulate their uptake and loss of gases by 'distributed computation' -- a kind of information processing that involves communication between many interacting units.

Nature adds this is similar to signals exchanged by ants to find the best source of food for an ant community.

> This might not sound much like what a computer does, but it is. In distributed computation, signals exchanged between components of the system define the process for solving a problem. Researchers are now exploring the possibility of using distributed computing with swarms of simple robots to carry out tasks, such as searching a landscape, more efficiently than a single, more sophisticated robot could manage.

Let's come back to plants and their leaves, which have tiny pores on their surface, called stomata.

Here is a picture of a cactus leaf showing these stomata, which permit the plant to breathe when they're opened (Credit: Link).

January 2004

Sun	Mon	Tue	Wed	Thu	Fri	Sat
				1	2	3
4	5	6	7	8	9	10
11	12	13	14	15	16	17
18	19	20	21	22	23	24
25	26	27	28	29	30	31

Dec Feb

Search this blog for

Find ALL words Search

Courtesy of PicoSearch

Personal Links

Roland's weblog
Weblog Archives
Sidebar Stories
Rate this Blog
Roland's website

Other Links

Ars Technica
BoingBoing
Daily Rotation News
Geek.com
Gizmodo
Microdoc News
Nanodot
Slashdot
Smart Mobs
Techdirt
Technorati

People

Dave Barry
Paul Boutin

The Computing Continuum

- Each strikes a different balance
  - computation/communication coupling
- Implications for execution efficiency
- *Applications for diverse needs*
  - *computing is only one part of the story!*

Grids vs. Capability vs. Cluster Computing

- Not an "either/or" question
  - Each addresses different needs
  - Each are part of an integrated solution
- Grid strengths
  - Coupling necessarily distributed resources
    - instruments, software, hardware, archives, and people
  - Eliminating time and space barriers
    - remote resource access and capacity computing
  - Grids are not a cheap substitute for capability HPC
- Capability computing strengths
  - Supporting foundational computations
    - terascale and petascale "nation scale" problems
  - Engaging tightly coupled computations and teams
- Clusters
  - Low cost, group solution
  - Potential hidden costs
- Key is easy access to resources in a transparent way

Real Crisis With HPC Is With The Software

- Programming is stuck
  - Arguably hasn't changed since the 60's
- It's time for a change
  - Complexity is rising dramatically
    - highly parallel and distributed systems
      - From 10 to 100 to 1000 to 10000 to 100000 of processors!!
    - multidisciplinary applications
- A supercomputer application and software are usually much more long-lived than a hardware
  - Hardware life typically five years at most.
  - Fortran and C are the main programming models
- Software is a major cost component of modern technologies.
  - The tradition in HPC system procurement is to assume that the software is free.
- We don't have many great ideas about how to solve this problem.

Collaborators / Support

- TOP500
  - H. Meuer, Mannheim U
  - H. Simon, NERSC
  - E. Strohmaier, NERSC