

New Computing Initiatives in the US

Horst D. Simon
Director, NERSC Center Division, LBNL
June 25, 2003
http://www.nersc.gov/~simon





Signposts of Change in HPC

In early 2002 there were several signposts, which signal a fundamental change in HPC in the US:

- Installation and very impressive early performance results of the Earth Simulator System (April 2002)
- Lack of progress in computer architecture research evident at Petaflops Workshop (WIMPS, Feb. 2002)
- Poor or non-existing benchmarks on sustained systems performance (SSP) for the NERSC workload (March 2002)



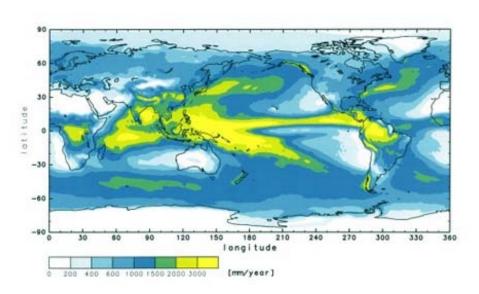
The Earth Simulation in Japan

Linpack benchmTF/s = 87% of 40

Completed Apri

Driven by climate earthquake simulation

Gordon Bell Prize at SC2002



http://www.es.jamstec.go.jp/esrdc/eng/menu.html

THE	-	沈み込み Subduction		
下降流 Descent style	diality.	2.2	巨大海台群 Ocean unit gr	roup
-	7	Part Part	8	
対流 Convecti	- 180 1	L. Harden		有太平洋 トットスポット群 South Pacific
	外核 Outer Core	(A)		hot spot group
内核 Inner Co	H.C.S.:r 期核元素 Parents nuclear el	ement _±#		中央海猫 Central

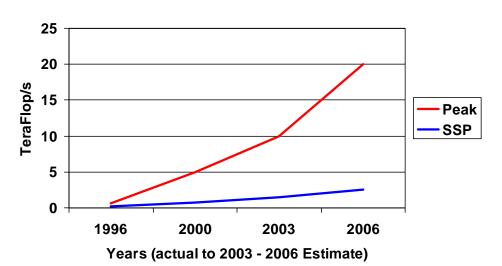
Understanding and Prediction of Global Climate Change	Understanding of Plate Tectonics
Occurrence prediction of meteorological disaster	Understanding of long- range crustal movements
Occurrence prediction of El Niño	Understanding of mechanism of seismicity
Understanding of effect of global warming	Understanding of migration of underground water and materials transfer in strata
Establishment of simulation technology with 1km resolution	



The Divergence Problem

- The requirements of high performance computing for science and engineering and the requirements of the commercial market are diverging.
- The commercial cluster of SMP approach is no longer sufficient to provide the highest level of performance
 - Lack of memory bandwidth
 - High interconnect latency
 - Lack of interconnect bandwidth
 - Lack of high performance parallel I/O
 - High cost of ownership for large scale systems

Divergence





Recent opinions on commodity technology in supercomputing



 "Gordon Bell, now a senior researcher at Microsoft, warns that off-the-shelf supercomputing is a dead end."

quoted from MIT Technology Review, Feb 2003.



"Beowulf is dead"

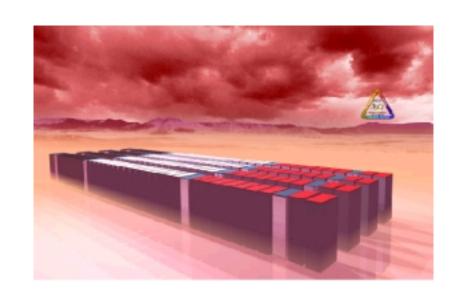
Thomas Sterling, Caltech, quoted from a panel discussion at SC2002, Nov. 2002



Cray-Sandia Cooperative Development: "Red Storm"

- Collaboration between Sandia Natl. Lab. and Cray (2004)
- True MPP, designed to be a single system
- Distributed memory MIMD parallel supercomputer
- Fully connected 3-D mesh interconnect. Each compute node processor has a bi-directional connection to the primary communication network.
- 108 compute node cabinets and 10,368 compute node processors (AMD Sledgehammer @ 2.0 GHz) ~20 Tflop/s peak
- ~10 TB of DDR memory @ 333 MHz
- 240 TB of disk storage (120 TB per color)
- Less than 2 MW total power and cooling.
- Less than 3,000 square feet of floor space

Courtesy: Bill Camp and Jim Thompkins, Sandia



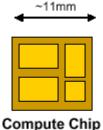


Blue Gene L

(2004: 180 TF system at LLNL)

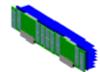
Building BlueGene/L

(compare this with a 1988 Cray YMP/8 at 2.7GF/s)



2 processors

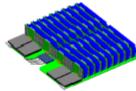
2.8/5.6 GF/s 4 MiB* eDRAM



Compute Card

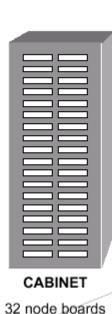
FRU 25mmx32mm 2 compute chips (2x1x1) 2.8/5.6 GF/s 256 MiB* DDR

15 W

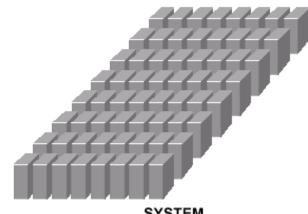


Node Board

32 compute chips 16 compute cards (4x4x2) 90/180 GE/s 8 GiB* DDR



(8x8x16) 2.9/5.7 TF/s 266 GiB* DDR 15-20 kW



SYSTEM

64 cabinets (32x32x64) 180/360 TF/s 16 TiB* ~1 MW 2500 sq.ft.

http://physics.nist.gov/cuu/Units/binary.html

*MiB = 2²⁰ bytes = 1,048,576 bytes ≈ 10⁶ + 5% bytes *GiB = 2³⁰ bytes = 1,073,741,824 bytes ≈ 10⁹ + 7% bytes *TiB = 2^{40} bytes = 1,099,511,627,776 bytes $\approx 10^{12} + 10\%$ bytes *PiB = 2^{60} bytes = 1,152,921,504,606,846,976 bytes $\approx 10^{15} + 15\%$ bytes

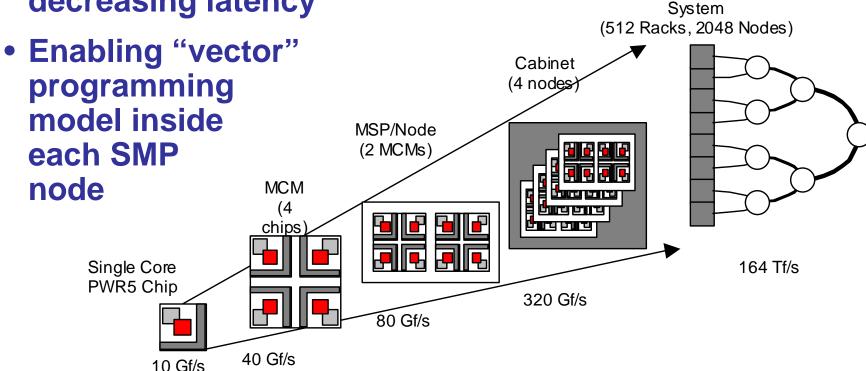


Blue Planet: A Conceptual View

(2005: NERSC and LLNL)

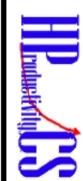
 Increasing memory bandwidth – single core chips with dedicated caches for 8 way nodes

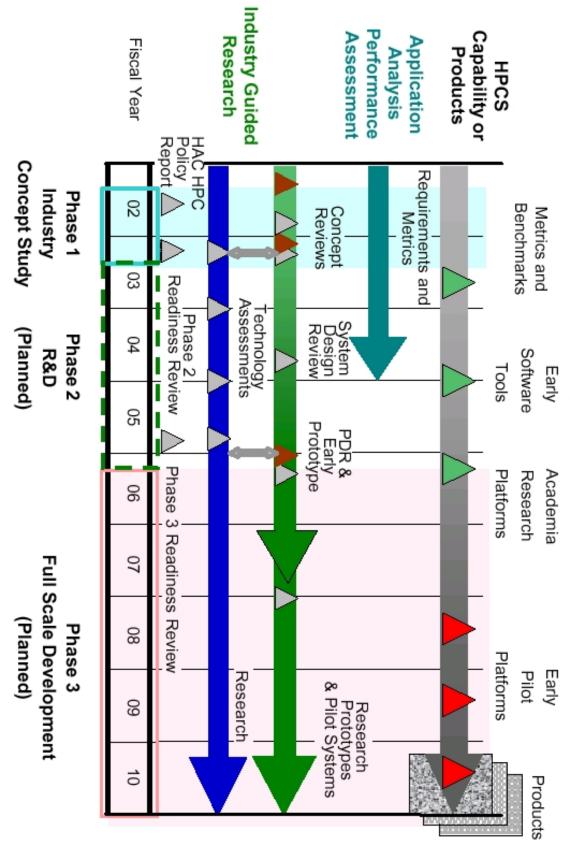
Increasing switch bandwidth and decreasing latency





HPCS Planned Program Phases 1-3







Industry BAA Open/Close

Critical Program Milestones



Renewed Interest in High Performance Computing at the Political Level

- HECRTF (High End Computing Revitalization Task Force)
 - OSTP charted; will produce roadmap for all federal agencies
- DOE/SC: Ultrascale Initiative (FY2004)
- DOE/NNSA: ASCI plans until 2015
- NSF: Cyberinfrastructure
- DOD: "IHEC" report
- NAS: study on "The Future of Supercomputing"