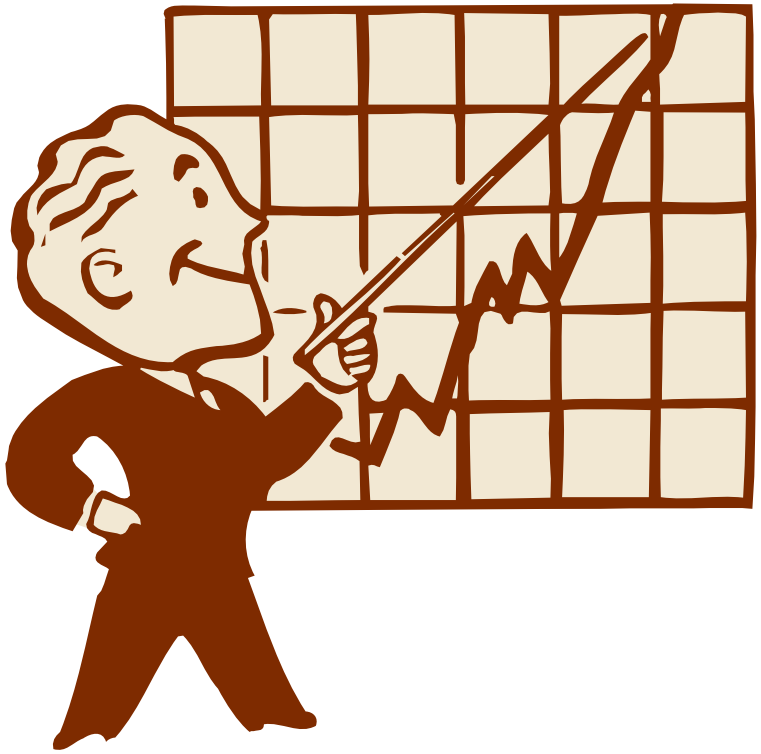
Scaling to Petaflops on Cray's MPP Supercomputer

John M. Levesque

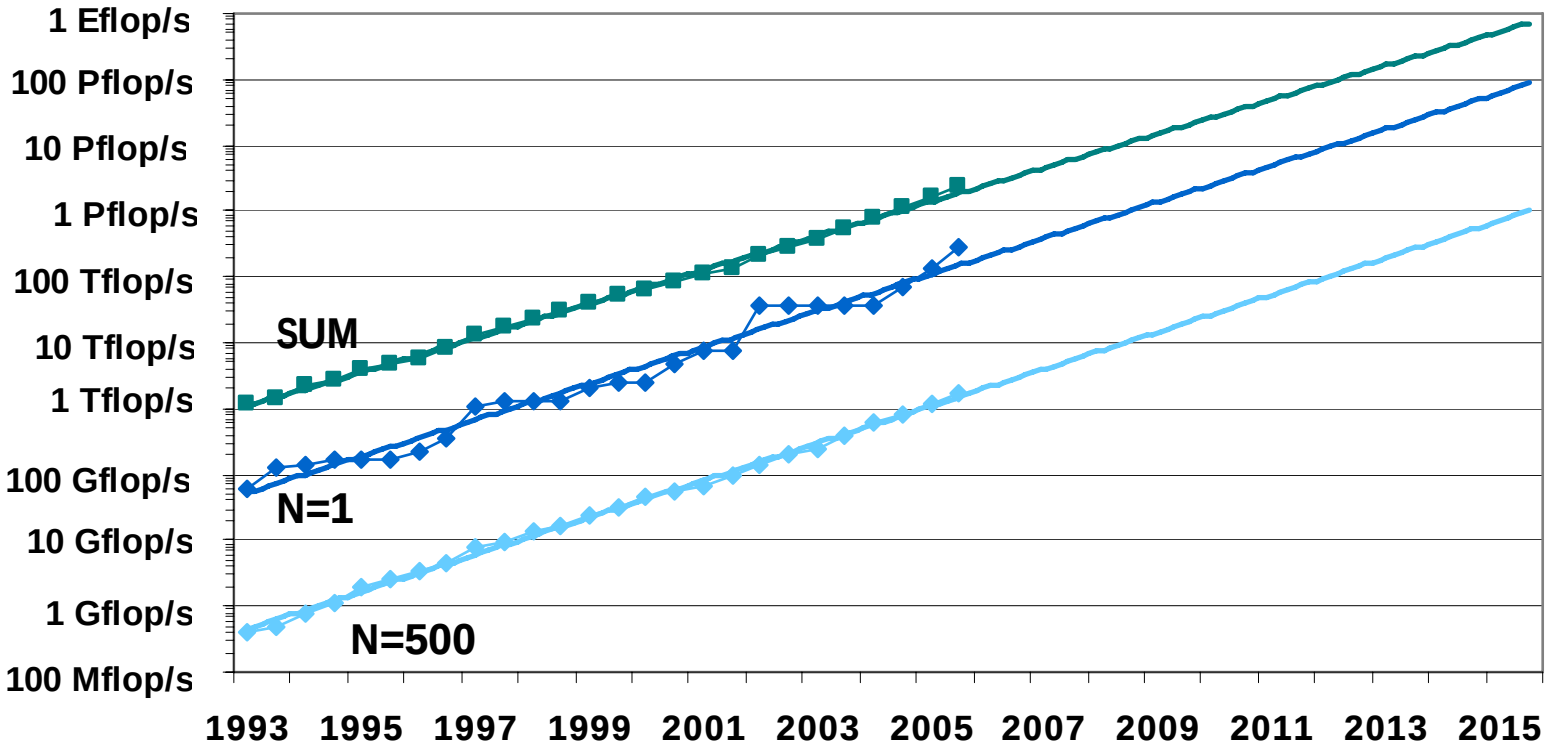
Director – Cray's Supercomputing Center of Excellence

Oct, 2006

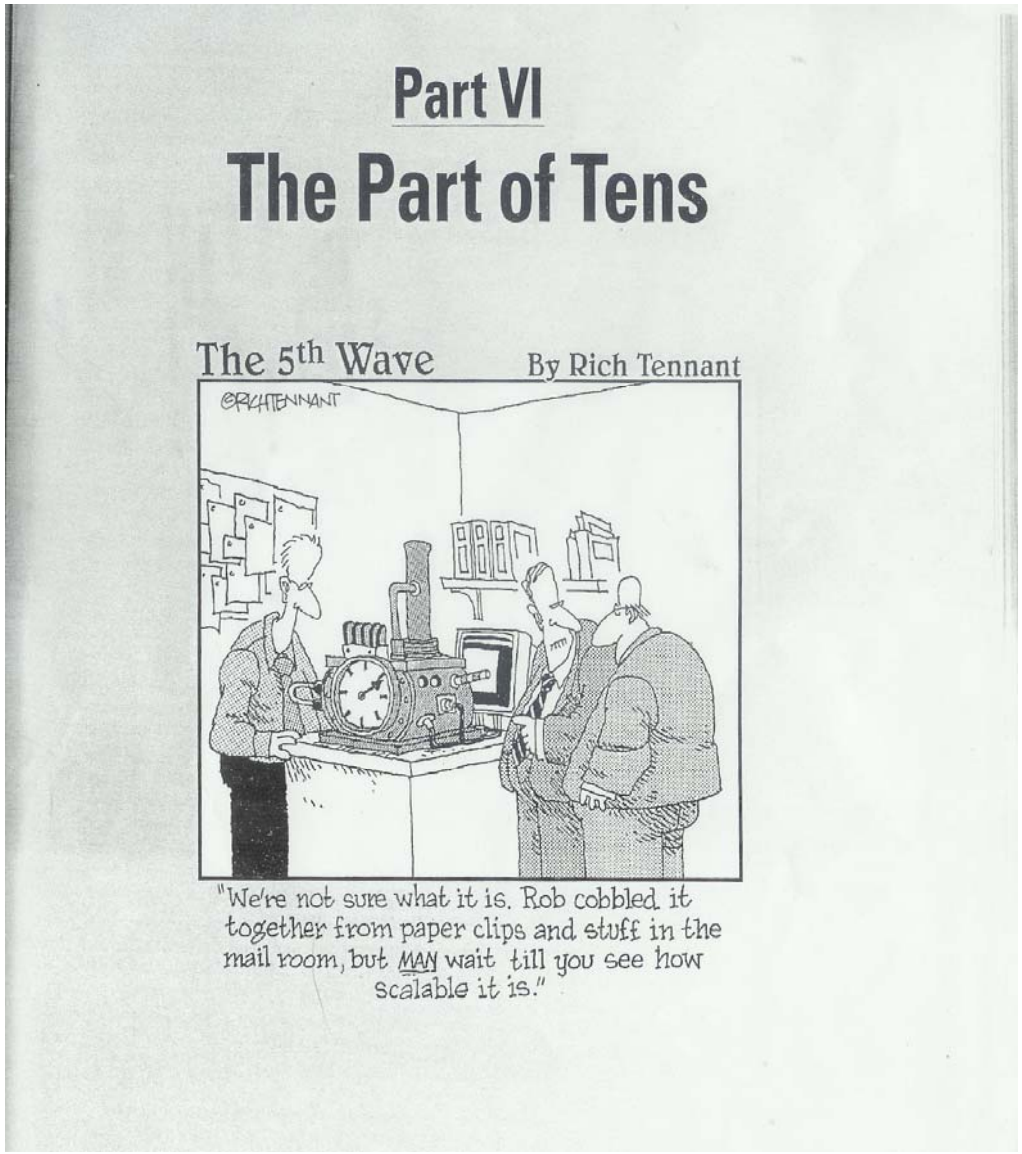
What Do you mean by Scaling to a Petaflop?



Well I don‘t mean Linpack!



The Cray XT Line



Part VI

The Part of Tens

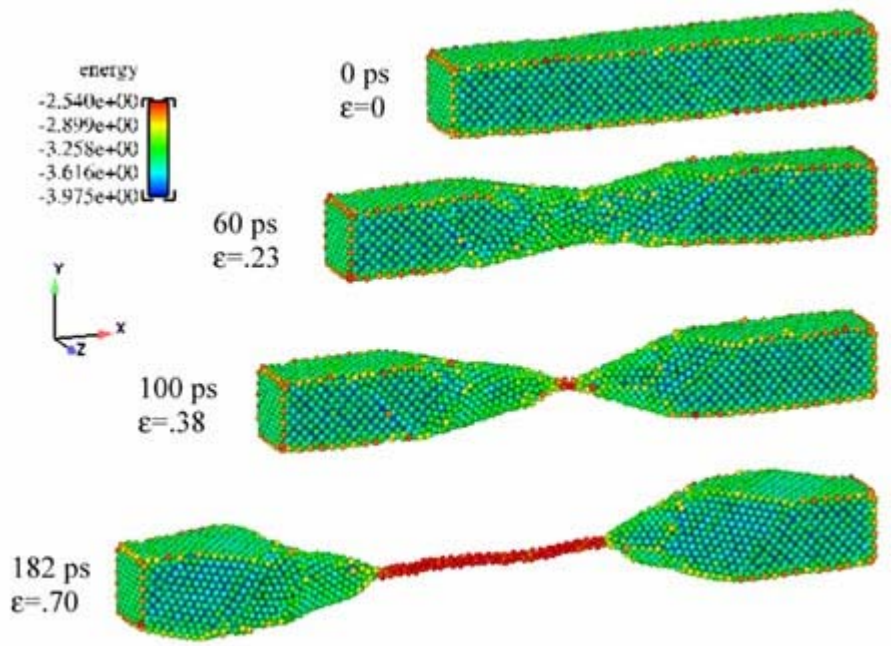
The 5th Wave By Rich Tennant

"We're not sure what it is. Rob cobbled it together from paper clips and stuff in the mail room, but MAN wait till you see how scalable it is."

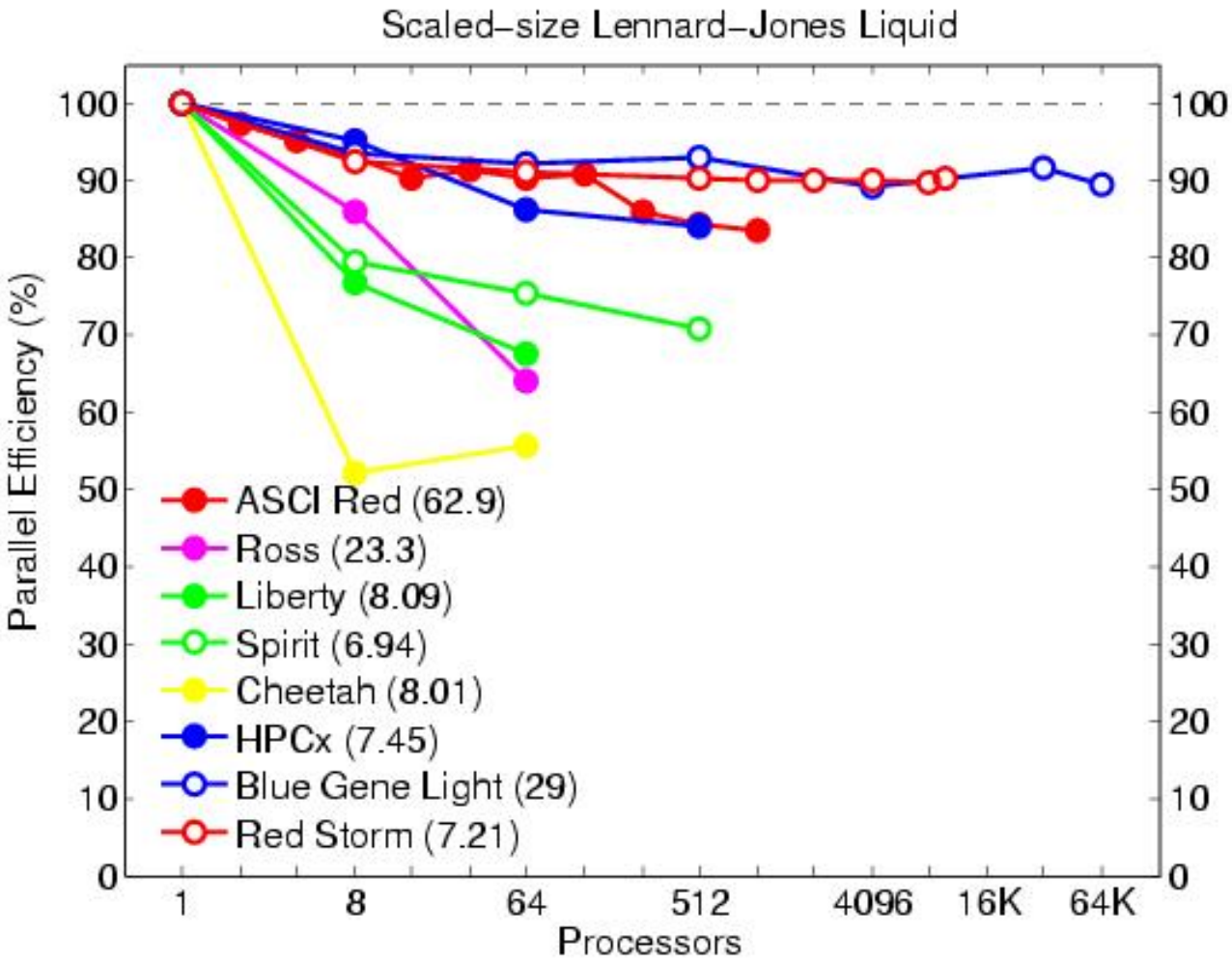
I mean ?

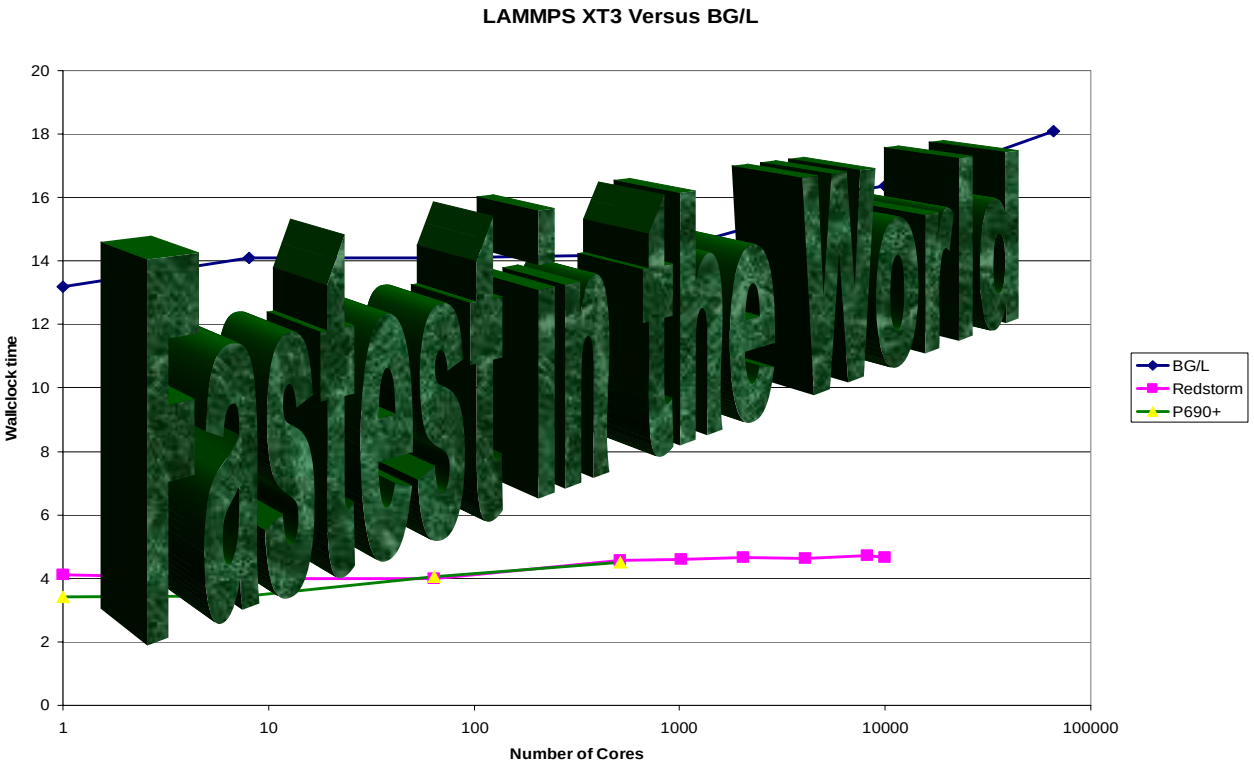
LAMMPS

Large-scale Atomic/Molecular Massively Parallel Simulator



Nickname	Vendor/ Machine	Processors	Site	CPU	Interconnect	Bandwidth	Latency
ASCI Red	Intel	1500	SNL	333 MHz Pentium III	custom	310	18
Ross	custom Linux cluster	64	SNL	500 MHz DEC Alpha	Myrinet	100	65
Liberty	HP Linux cluster	64	SNL	3.0 GHz dual Xeons (32-bit)	Myrinet	230	9
Spirit	HP Linux cluster	512	SNL	3.4 GHz dual Xeons (64-bit)	Myrinet	230	9
Cheetah	IBM p690	64	ORNL	1.3 GHz Power4	custom	1490	7
HPCx	IBM p690+	512	Daresbury	1.7 GHz Power4+	custom	1450	6
Blue Gene Light	IBM	65536	LLNL	700 MHz PowerPC 440	custom	150	3
Red Storm	Cray XT3	10000	SNL	2.0 GHz Opteron	custom	1100	7


Scaled–size Lennard–Jones Liquid
Parallel Efficiency (%)
Processors
1
8
64
512
4096
16K
64K
0
10
20
30
40
50
60
70
80
90
100
ASCI Red (62.9)
Ross (23.3)
Liberty (8.09)
Spirit (6.94)
Cheetah (8.01)
HPCx (7.45)
Blue Gene Light (29)
Red Storm (7.21)

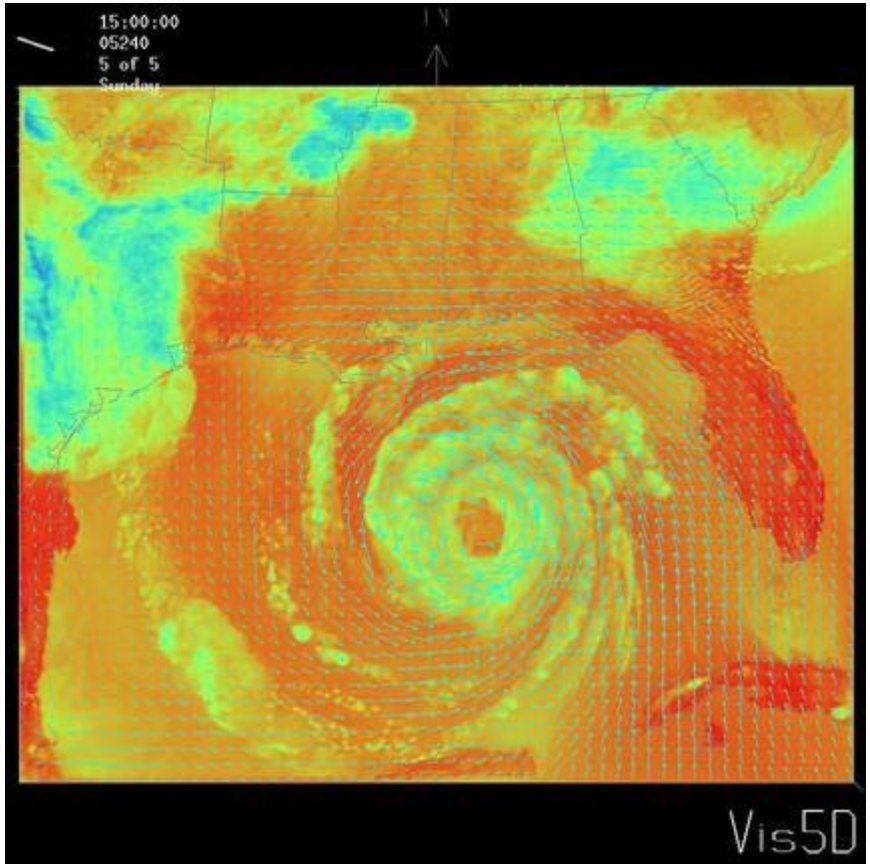


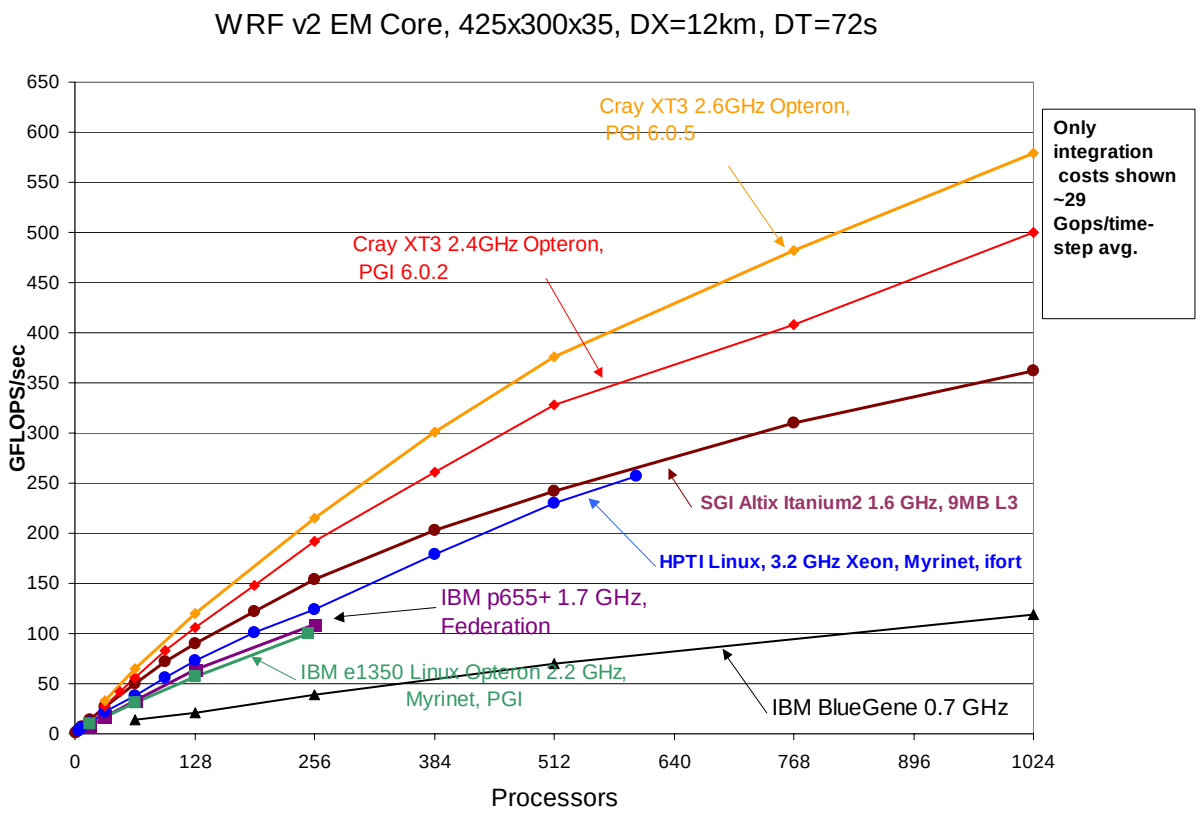
Latest XT3's are 2.6 GHZ

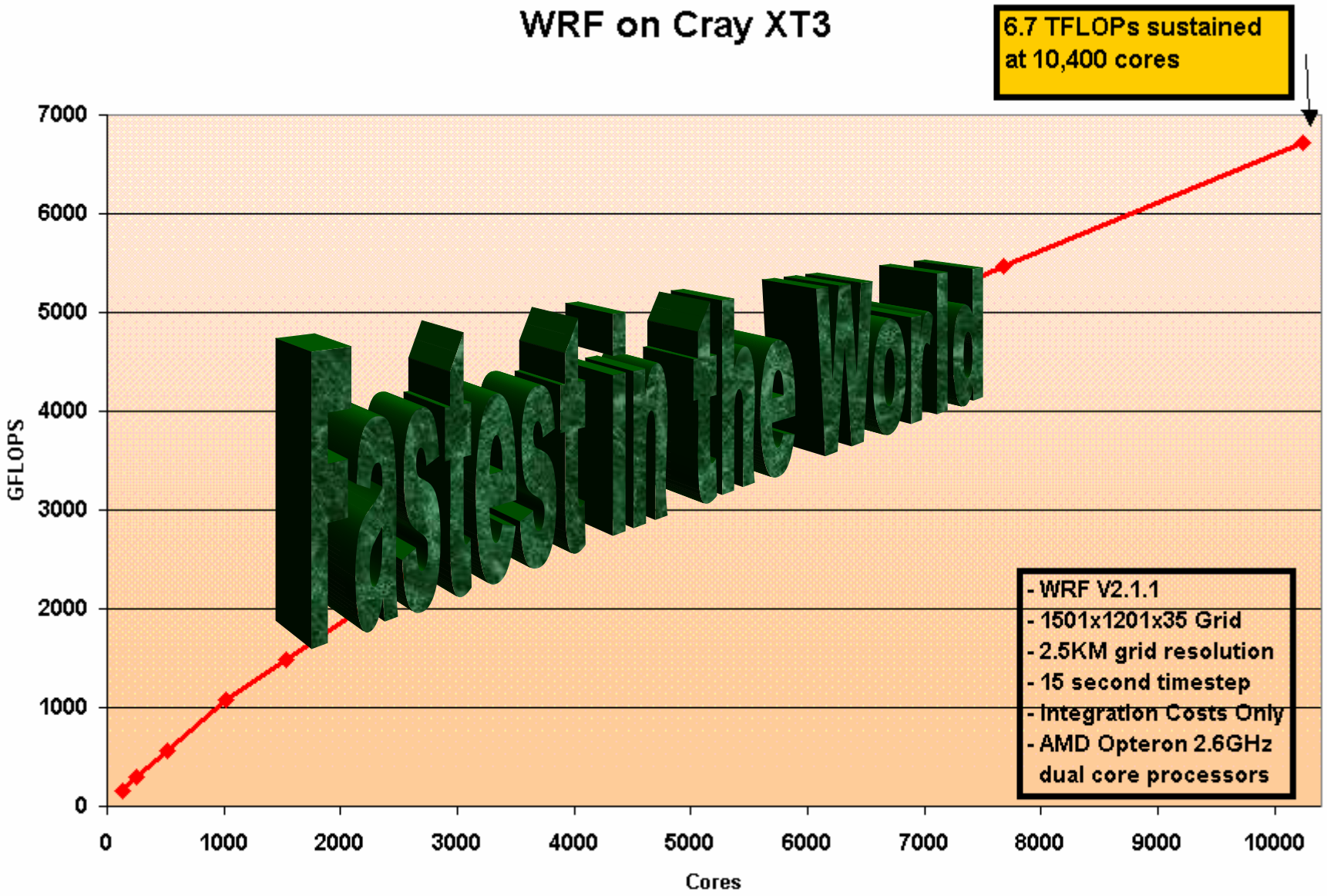
I mean ?

WRF

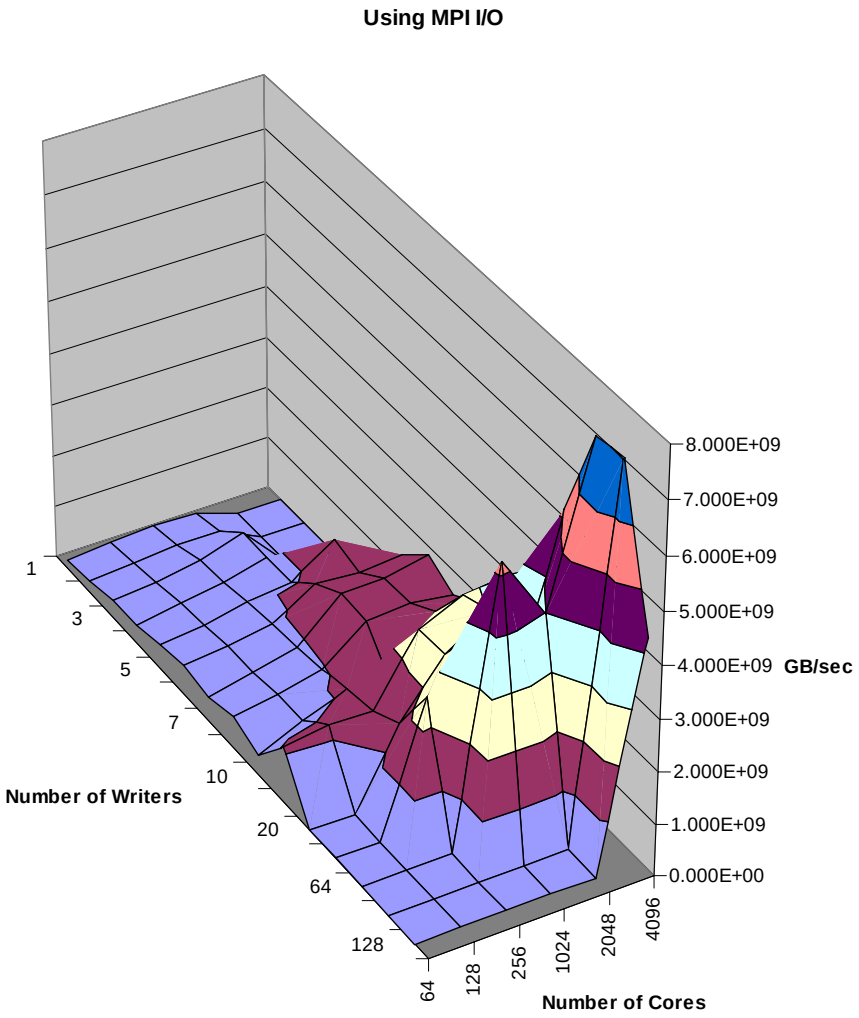
Weather Research and Forecast Model




WRF v2 EM Core, 425x300x35, DX=12km, DT=72s
GFLOPS/sec
Processors
0
50
100
150
200
250
300
350
400
450
500
550
600
650
0
128
256
384
512
640
768
896
1024
Cray XT3 2.6GHz Opteron, PGI 6.0.5
Cray XT3 2.4GHz Opteron, PGI 6.0.2
SGI Altix Itanium2 1.6 GHz, 9MB L3
HPTI Linux, 3.2 GHz Xeon, Myrinet, ifort
IBM p655+ 1.7 GHz, Federation
IBM e1350 Linux Opteron 2.2 GHz, Myrinet, PGI
IBM BlueGene 0.7 GHz
Only integration costs shown ~29 Gops/time-step avg.


WRF on Cray XT3
6.7 TFLOPs sustained at 10,400 cores
Fastest in the World
- WRF V2.1.1
- 1501x1201x35 Grid
- 2.5KM grid resolution
- 15 second timestep
- Integration Costs Only
- AMD Opteron 2.6GHz dual core processors
GFLOPS
0
1000
2000
3000
4000
5000
6000
7000
Cores
0
1000
2000
3000
4000
5000
6000
7000
8000
9000
10000

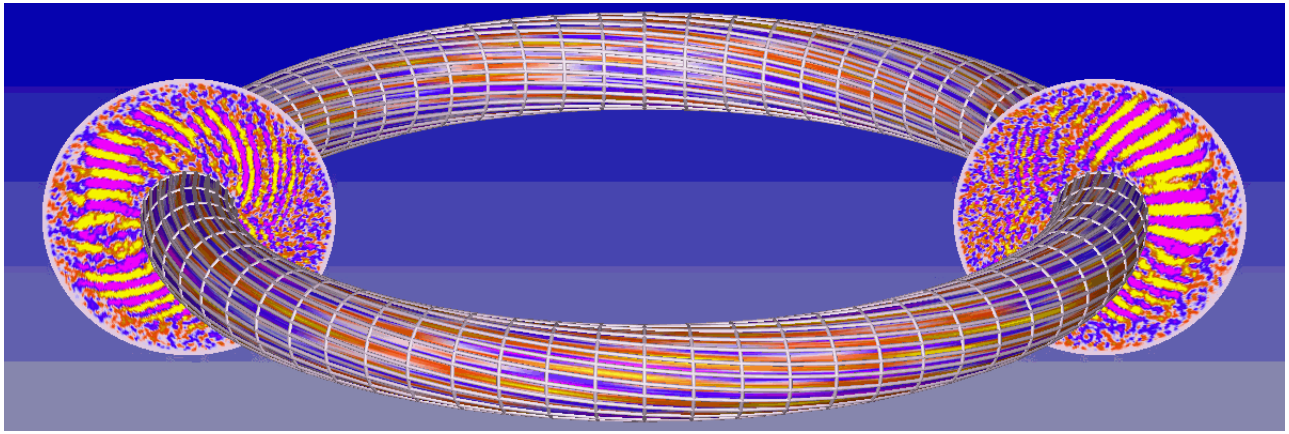
What about I/O?



I mean ?

GTC

Gyrokinetic Particle Simulation of Turbulent Transport in Burning Plasmas



News Briefs:

ORNL's Jaguar Fastest for Fusion Turbulence Simulations

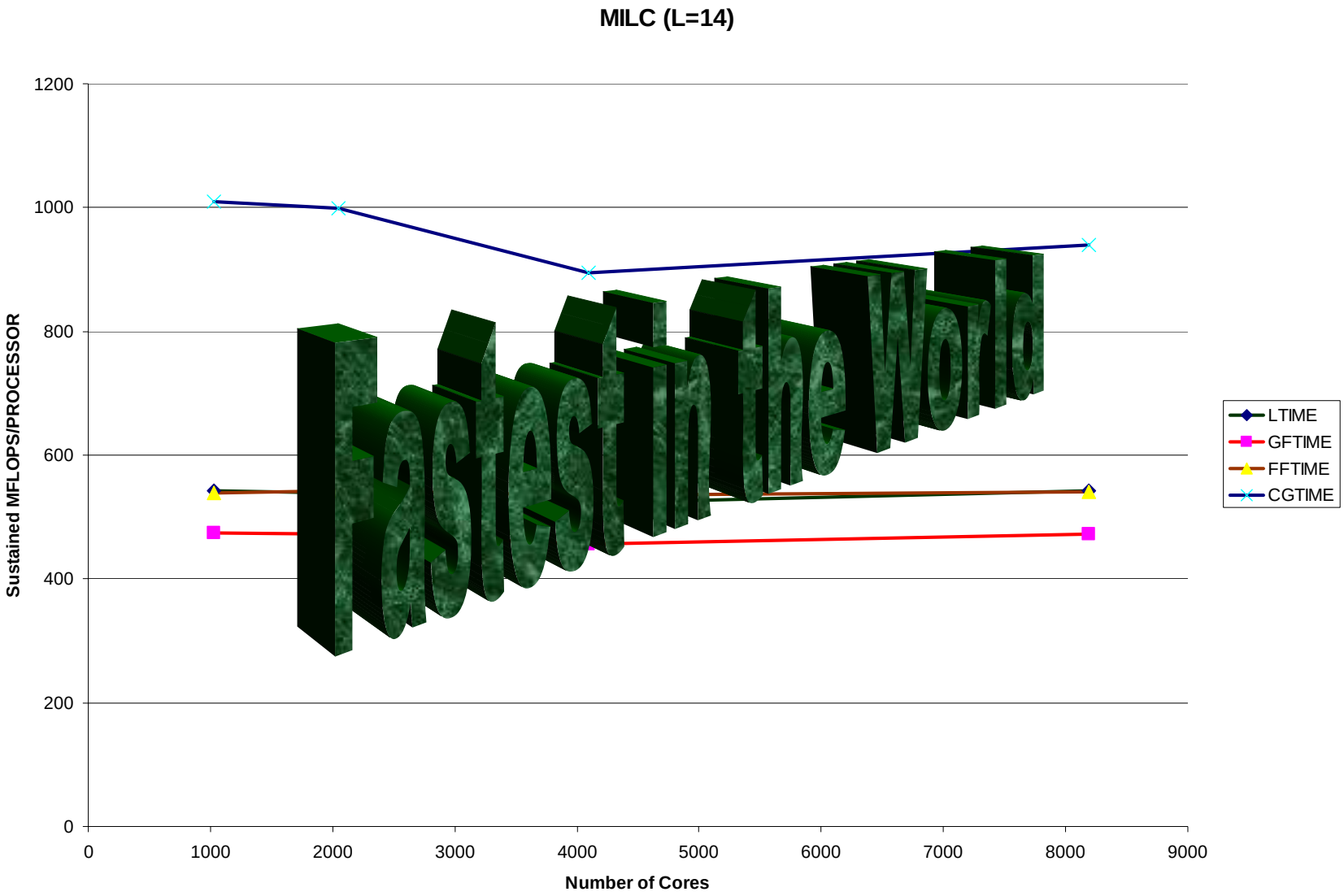
http://www.hpcwire.com/hpc/943289.html

Oak Ridge National Laboratory's (ORNL's) Cray XT3 supercomputer, known as Jaguar, has become the fastest system in the world for running the Princeton Plasma Physics Laboratory's (PPPL's) flagship code for studying plasma microturbulence in fusion reactors. PPPL's Stephane Ethier recently succeeded in running the Gyrokinetic Toroidal Code (GTC) on 10,386 of Jaguar's 10,424 processing cores, advancing 5.4 billion particles per step per second. That performance is a 13 percent improvement over the previous record of 4.8 billion particles per step per second set on Japan's Earth Simulator.

I mean ?

MILC

MIMD Lattice Computation



Run by Steve Gottlieb on ORNL's Jaguar System

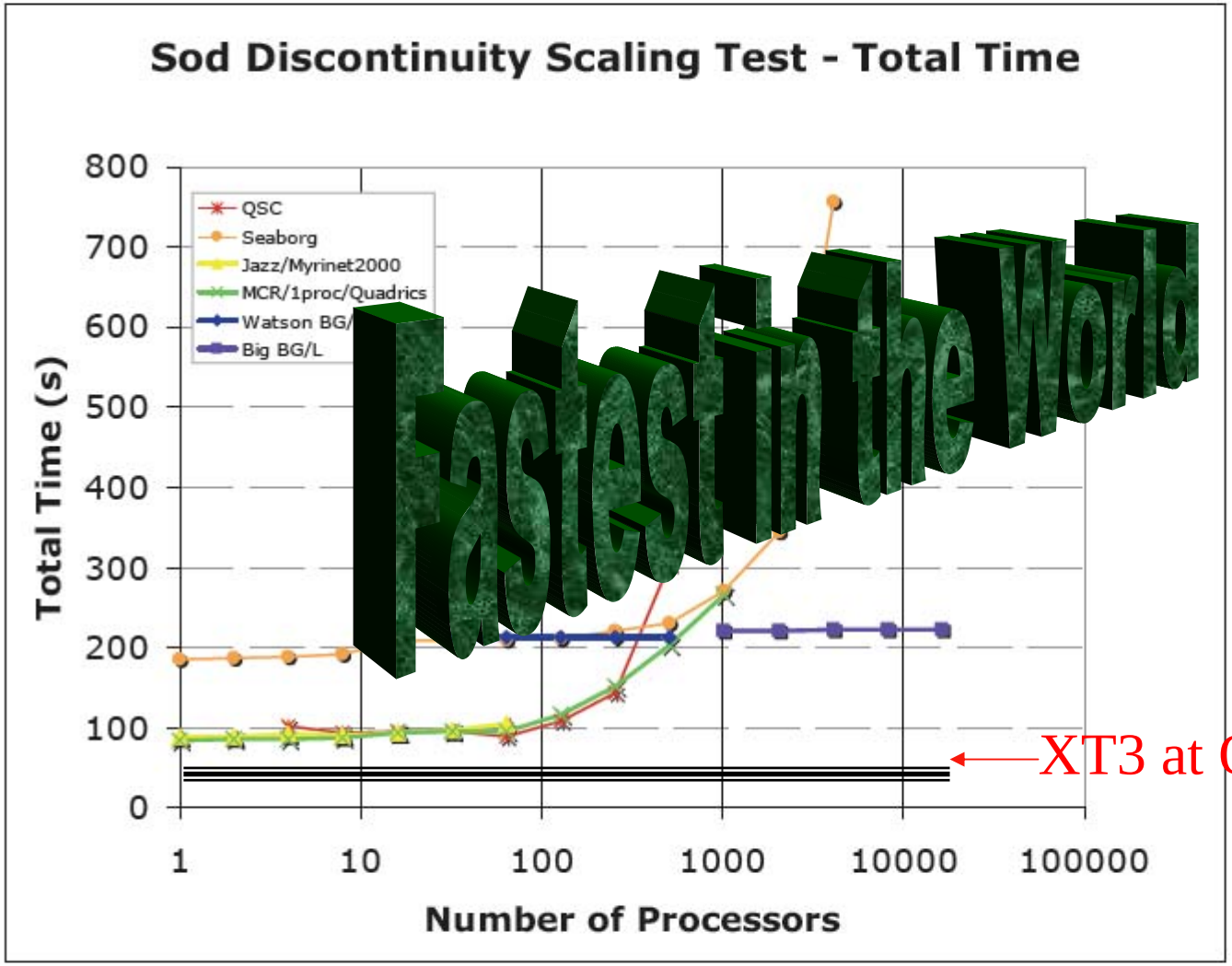
I mean ?

FLASH

State-of-the-art simulator code for solving nuclear astrophysical problems related to exploding stars.

FLASH

(Katherine Riley, Andrew Siegel - U Chicago, AN



QSC: 256nodex4way

Seaborg: IBM SP, 1.5
NERSC

Jazz : 350node, 2.4GH

MCR : 1152node,2.4G

Big BGL: 16 Racks, c

I mean ?

POP and CICE

Parallel Ocean Program and ICE Model

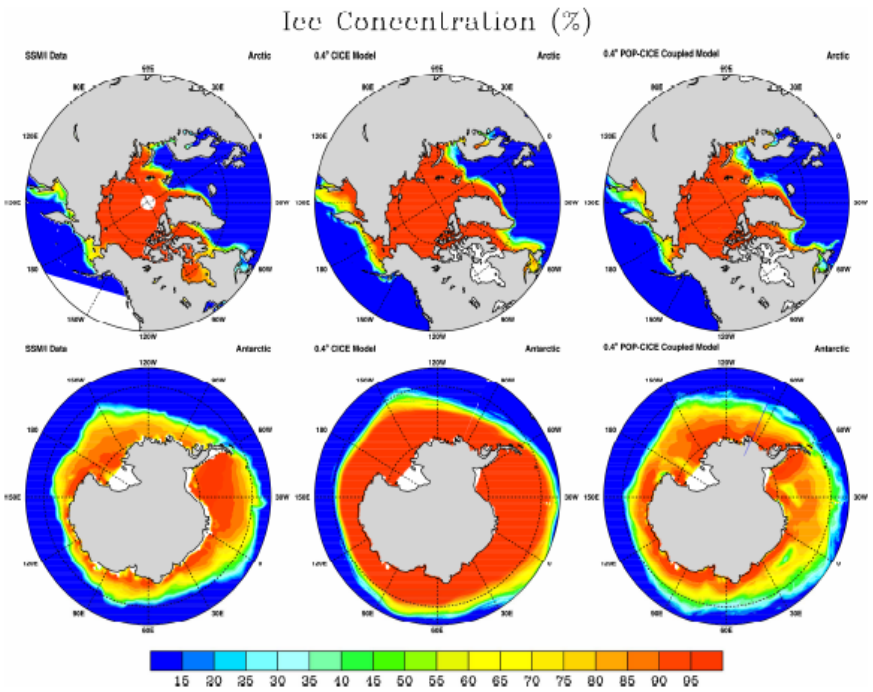
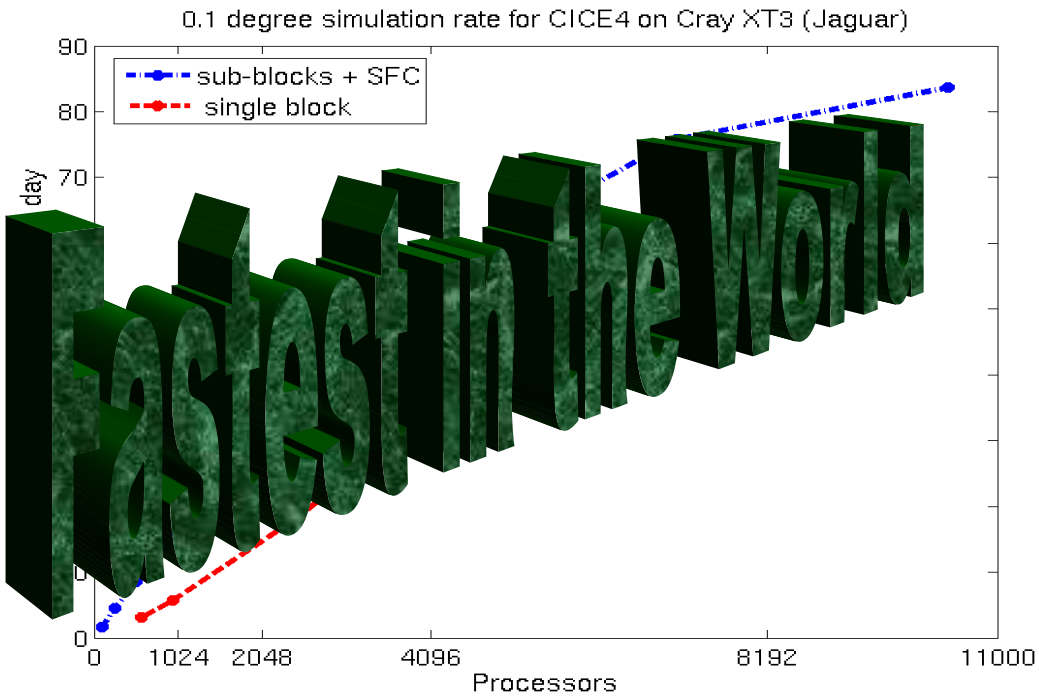
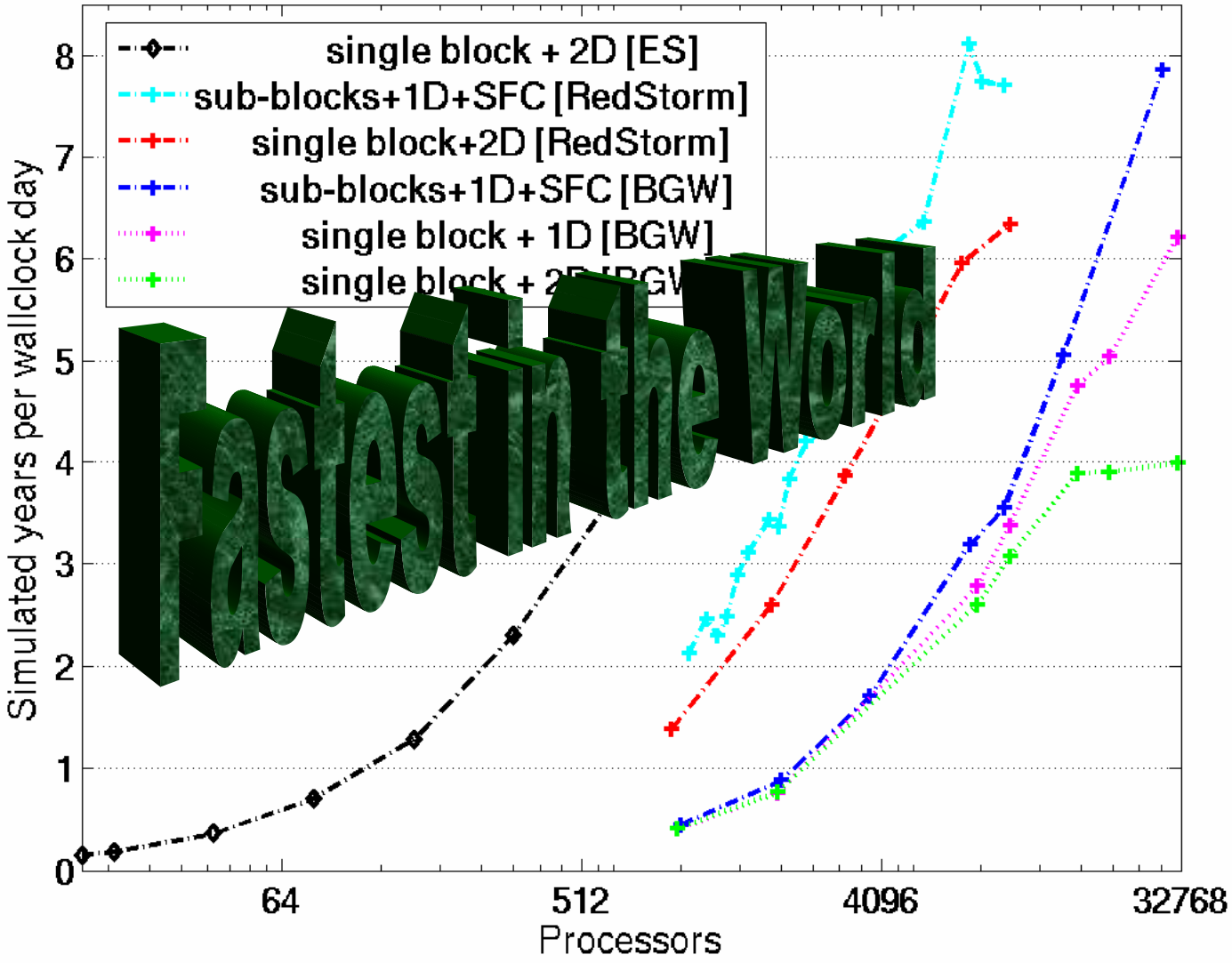


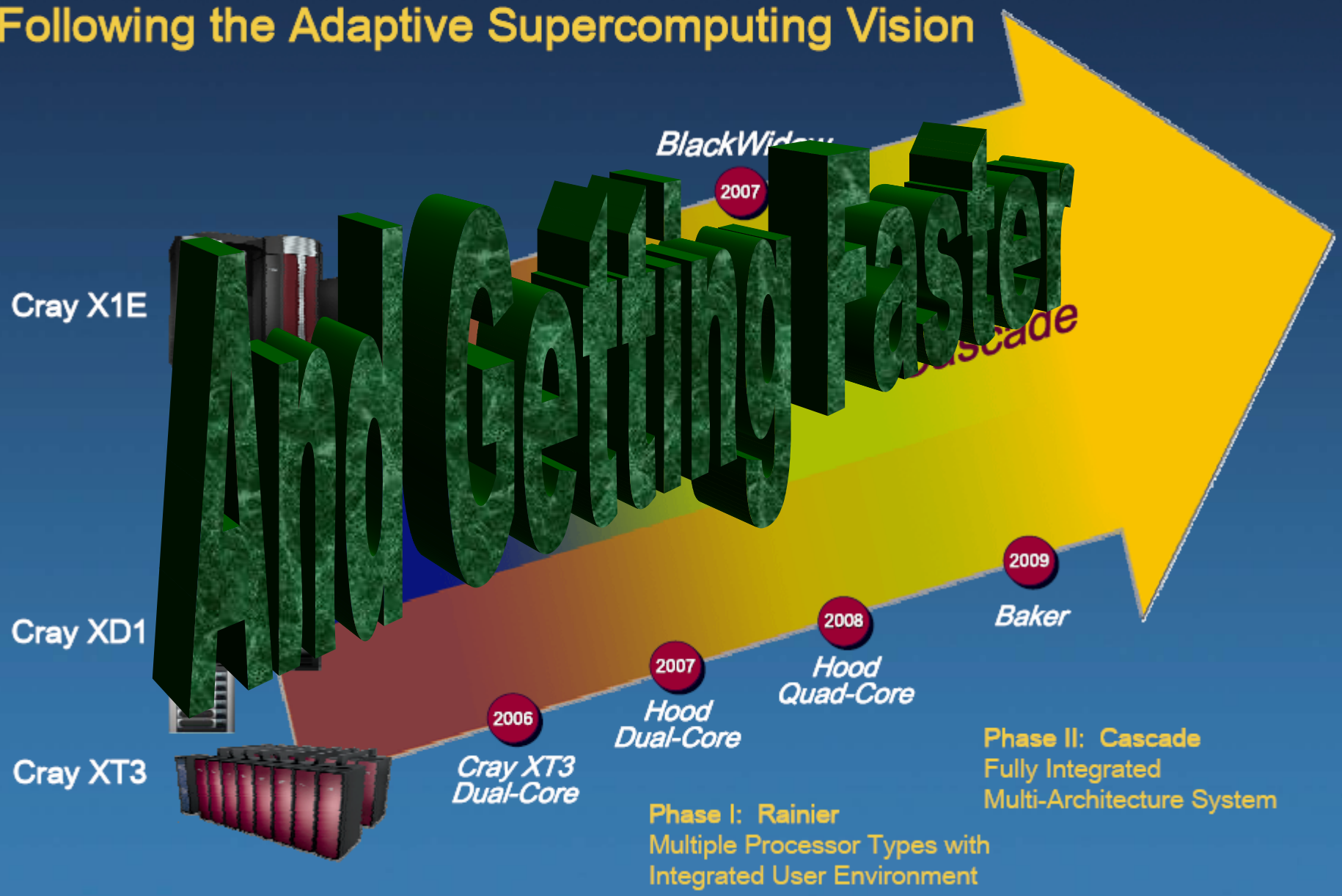
Figure 3: Sea ice concentration (%) in the Arctic (top) and Antarctic (bottom) from passive microwave data (left), and the 0.4° global stand-alone CICE (middle) and coupled POP/CICE (right) simulations for winter (1991-2001).


0.1 degree simulation rate for CICE4 on Cray XT3 (Jaguar)
sub-blocks + SFC
single block
day
90
80
70
0
0
1024
2048
4096
8192
11000
Processors
Fastest in the World


single block + 2D [ES]
sub-blocks+1D+SFC [RedStorm]
single block+2D [RedStorm]
sub-blocks+1D+SFC [BGW]
single block + 1D [BGW]
Simulated years per wallclock day
0
1
2
3
4
5
6
7
8
64
512
4096
32768
Processors
Fastest in the World

The Cray Roadmap

Following the Adaptive Supercomputing Vision



And Getting Faster