High-Performance Computing

at Oak Ridge National Lab

Dr. Olaf O. Storaasli
Future Technologies Group
Computer Science & Mathematics Division
Oak Ridge National Laboratory

Concordia College

15 Oct '09

OAK RIDGE NATIONAL LABORATORY
MANAGED BY UT-BATTELLE
FOR U.S. DEPARTMENT OF ENERGY

ORNL “X-10” History

1st Graphite Plutonium Reactor => PNL

ORNL: *DOE's #1 Energy & Science Lab, #1 Materials*

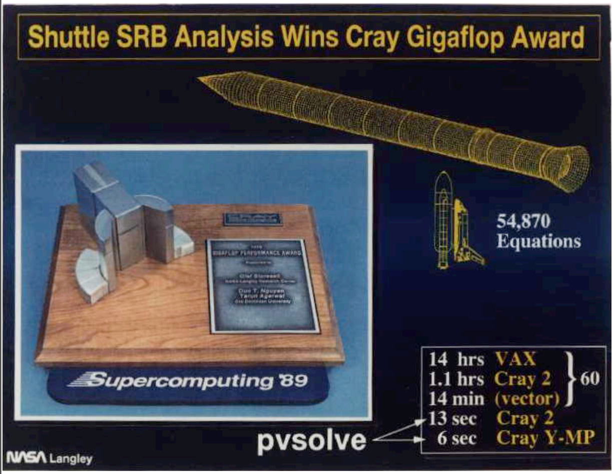
- 4K employees + 3K guest researchers
- #1 Science Supercomputers: DOE+NSF
- $1.3B+ SNS

See YouTube Video at:

http://www.youtube.com/watch?v=N7gqaHwSxcg

HPC Speedup: 1000X per decade

Evolution of the fastest sustained performance
in real simulations



Shuttle SRB Analysis Wins Cray Gigaflop Award

54,870
Equations

Supercomputing '89

14 hrs VAX
1.1 hrs Cray 2
14 min (vector) } 60
13 sec Cray 2
6 sec Cray Y-MP

pvsolve

NASA Langley

1.5 Gigaflop/s
Cray YMP
0.8 10^1 processors
(Storaasli - NASA)

1.02 Teraflop/s
Cray T_{3D}
1.5 10^3 processors
(ORNL)

1.35 Petaflop/s
Cray XT5
1.5 10^5 processor cores
(Shultness - ORNL)

~1 Exaflop/s
~10^7 processing units
(?)

1989 1998 2008 2018

1 Exaflop = 1,000 Petaflops = 1,000,000 Teraflops = 1,000,000,000 Gigaflops = 1,000,000,000,000 Mflops

Million-fold increase in computing and data capabilities

2004

Cray X1
3 TF

2005

Cray XT3
Single-core
26 TF

2006

Cray XT3
Dual-core
54 TF

2007

Cray XT4
119 TF

Cray XT4
Quad-core
166 TF
18 TB, 0.3 PB

2008

Cray XT4
Quad-core
263 TF
62 TB, 1 PB

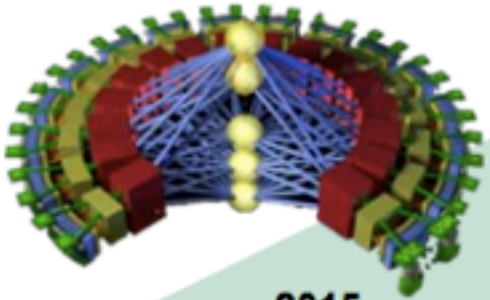
Cray "Baker"
8-core, dual-socket SMP
1.4 PF
300 TB, 10 PB

2009

Cray "Baker"
8/12-core, dual-socket SMP
~1 PF
100 TB, 2.5 PB

2011

DARPA
HPCS
20 PF

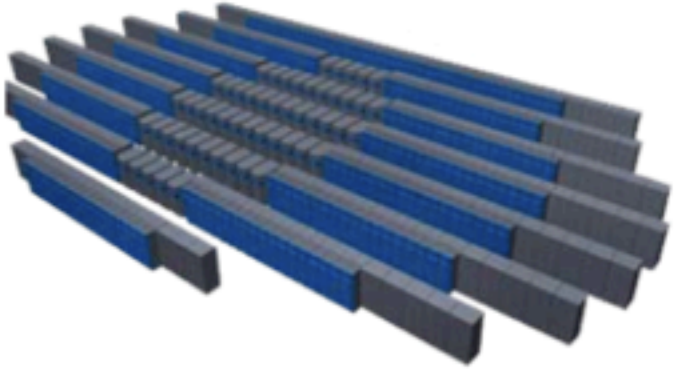


2015

Future
system
100–250 PF

2018

Future
system
1000 PF
(1 EF)



Slide courtesy of Thomas Zacharia

Jaguar: World's most powerful computer
Designed for science from the ground up

Peak performance	2.3 PetaFLOPS
System memory	362 terabytes
Disk space	10.7 petabytes
Disk bandwidth	240+ gigabytes/second
Interconnect bandwidth	532 terabytes/second

Slide courtesy of Thomas Zacharia

Cray XT5 portion of Jaguar @ NCCS

2.3 PetaFLOPS
6-core AMD
224,256 cores
2.3 GHz
200 cabinets
362TB memory
Details: nccs.gov

Kraken
World's most powerful academic computer

Peak performance	0.615 petaflops, 0.967 PF in late 2009
System memory	100 terabytes
Disk space	3.3 petabytes (raw)
Disk bandwidth	30 gigabytes/second
Interconnect bandwidth	532 terabytes/second

Slide courtesy of Thomas Zacharia

Oak Ridge National Laboratory to get 3rd supercomputer

Machine part of $215M research deal with NOAA

By Frank Munger

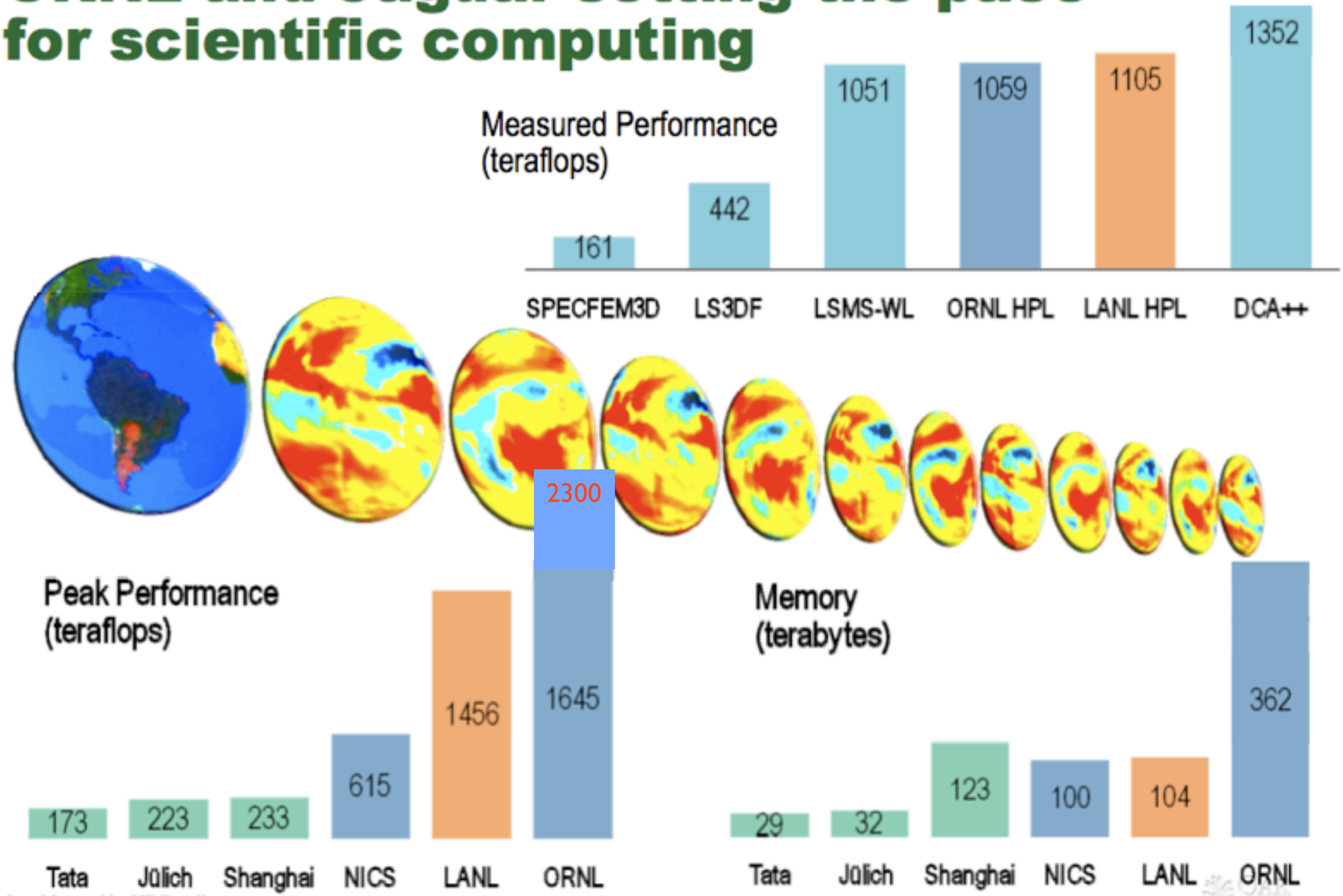
Thursday, September 24, 2009

OAK RIDGE - As part of its new five-year, $215 million climate research agreement with the National Oceanic and Atmospheric Administration, Oak Ridge National Laboratory will be acquiring yet another supercomputer.

The procurement process for the new machine is in the works, and, by this time next year, ORNL should have three computers capable of **at least one petaflops** (1,000 trillion calculations per second), according to Jeff Nichols, ORNL's interim computing chief.

"It'll be in the **same class as Jaguar and Kraken**," Nichols said, referring to the two Cray XT5 systems already housed in the lab's National Center for Computational Sciences.

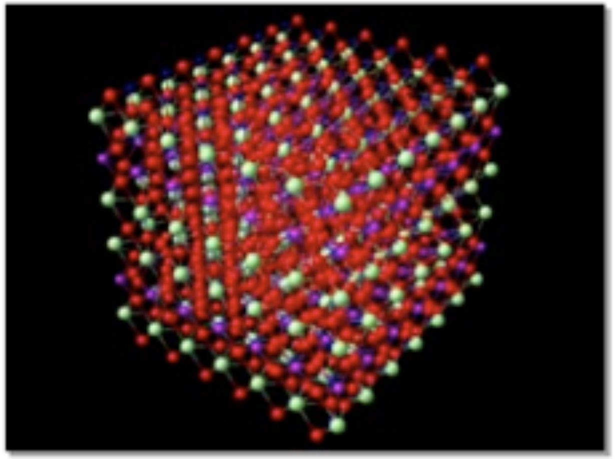
ORNL and Jaguar setting the pace for scientific computing



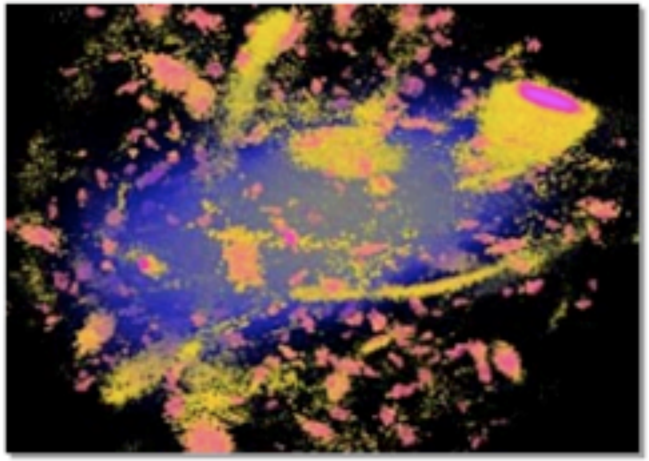
Slide courtesy of Thomas Zacharia

Enabling breakthrough science

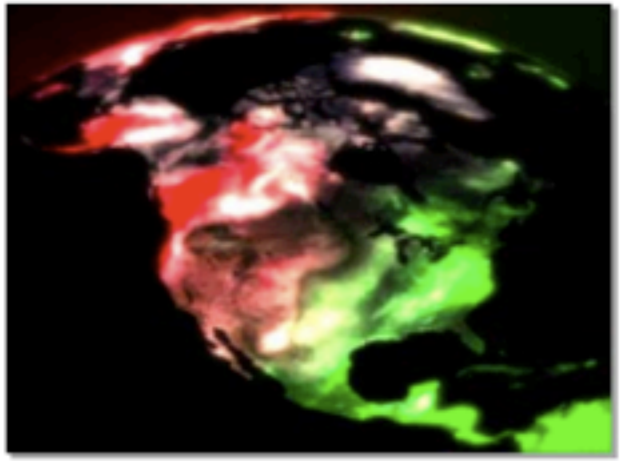
5 of top 10 ASCR science accomplishments in the past 18 months used LCF resources and staff



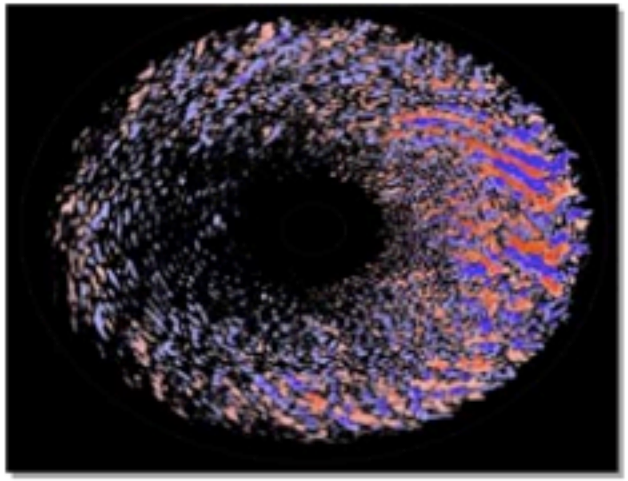
Electron pairing in HTSC cuprates
PRL (2007, 2008)



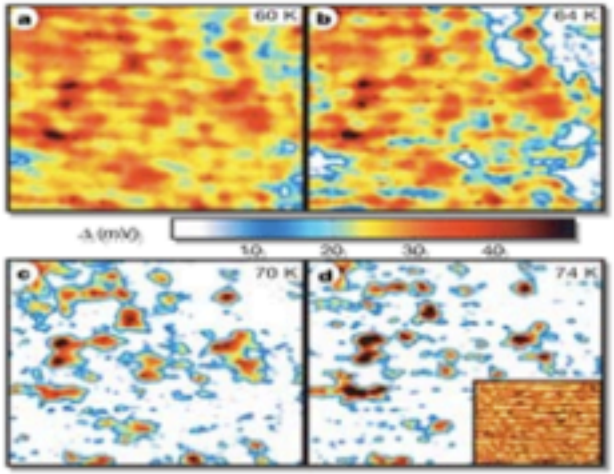
Shining a light on dark matter
Nature **454**, 735 (2008)



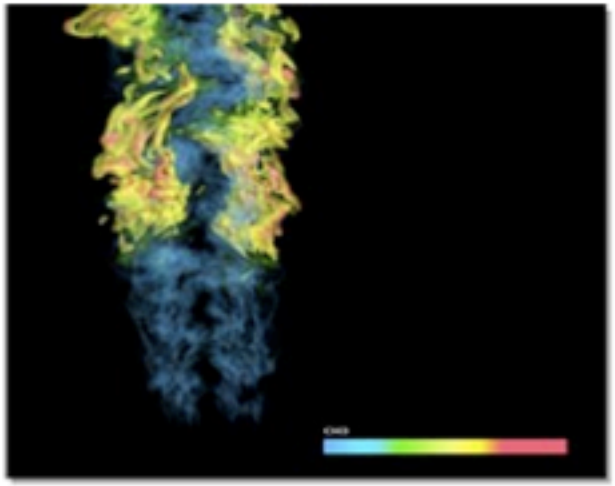
Modeling the full earth system



Fusion: Taming turbulent heat loss
PRL **99**, *Phys. Plasmas* **14**



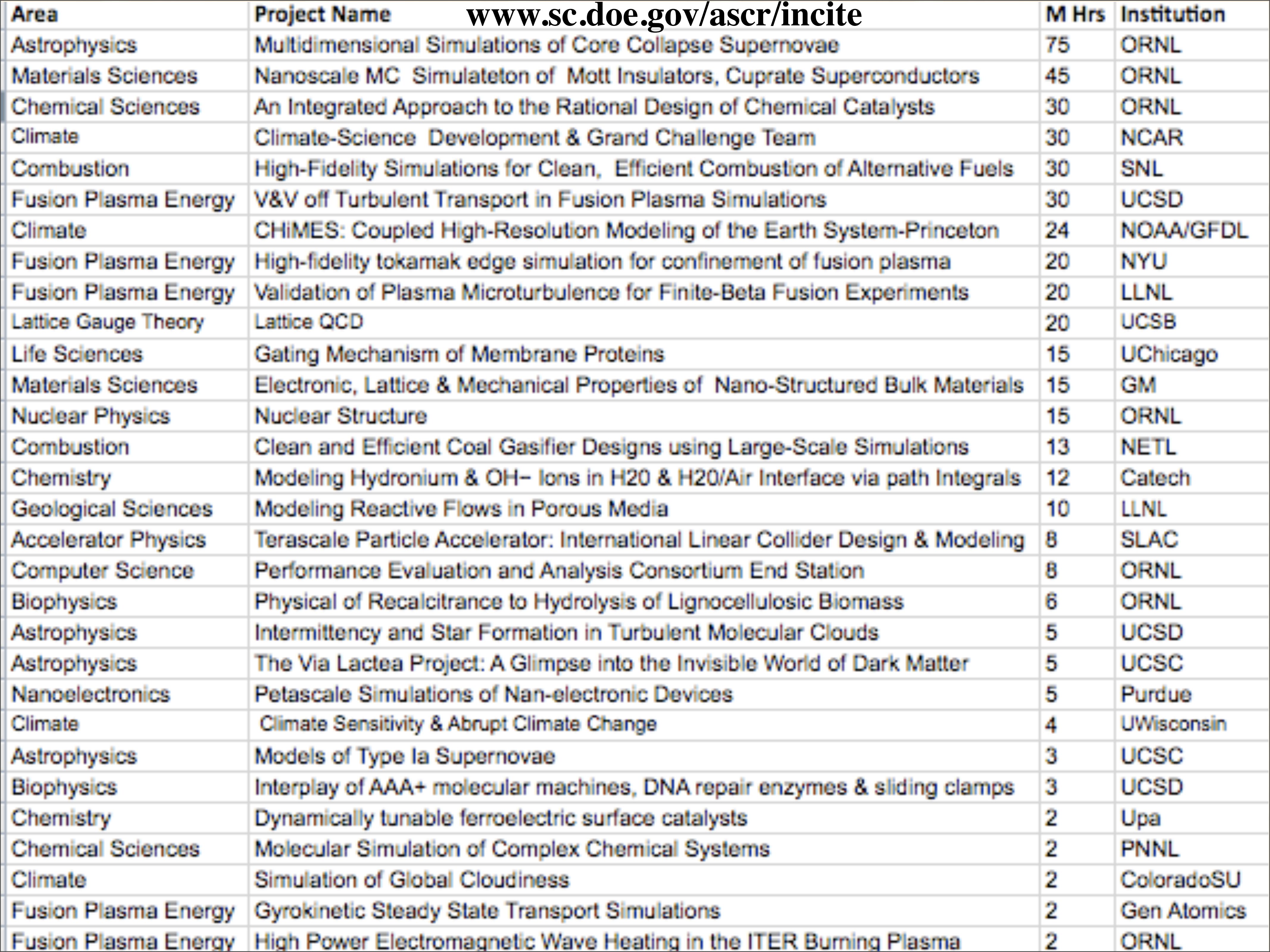
Nanoscale nonhomogeneities in high-temperature superconductors
Winner of Gordon Bell prize



Stabilizing a lifted flame
Combust. Flame (2008)

Slide courtesy of Thomas Zacharia

www.sc.doe.gov/ascr/incite



Area	Project Name	M Hrs	Institution
Astrophysics	Multidimensional Simulations of Core Collapse Supernovae	75	ORNL
Materials Sciences	Nanoscale MC Simulateton of Mott Insulators, Cuprate Superconductors	45	ORNL
Chemical Sciences	An Integrated Approach to the Rational Design of Chemical Catalysts	30	ORNL
Climate	Climate-Science Development & Grand Challenge Team	30	NCAR
Combustion	High-Fidelity Simulations for Clean, Efficient Combustion of Alternative Fuels	30	SNL
Fusion Plasma Energy	V&V off Turbulent Transport in Fusion Plasma Simulations	30	UCSD
Climate	CHiMES: Coupled High-Resolution Modeling of the Earth System-Princeton	24	NOAA/GFDL
Fusion Plasma Energy	High-fidelity tokamak edge simulation for confinement of fusion plasma	20	NYU
Fusion Plasma Energy	Validation of Plasma Microturbulence for Finite-Beta Fusion Experiments	20	LLNL
Lattice Gauge Theory	Lattice QCD	20	UCSB
Life Sciences	Gating Mechanism of Membrane Proteins	15	UChicago
Materials Sciences	Electronic, Lattice & Mechanical Properties of Nano-Structured Bulk Materials	15	GM
Nuclear Physics	Nuclear Structure	15	ORNL
Combustion	Clean and Efficient Coal Gasifier Designs using Large-Scale Simulations	13	NETL
Chemistry	Modeling Hydronium & OH− Ions in H20 & H20/Air Interface via path Integrals	12	Catech
Geological Sciences	Modeling Reactive Flows in Porous Media	10	LLNL
Accelerator Physics	Terascale Particle Accelerator: International Linear Collider Design & Modeling	8	SLAC
Computer Science	Performance Evaluation and Analysis Consortium End Station	8	ORNL
Biophysics	Physical of Recalcitrance to Hydrolysis of Lignocellulosic Biomass	6	ORNL
Astrophysics	Intermittency and Star Formation in Turbulent Molecular Clouds	5	UCSD
Astrophysics	The Via Lactea Project: A Glimpse into the Invisible World of Dark Matter	5	UCSC
Nanoelectronics	Petascale Simulations of Nan-electronic Devices	5	Purdue
Climate	Climate Sensitivity & Abrupt Climate Change	4	UWisconsin
Astrophysics	Models of Type Ia Supernovae	3	UCSC
Biophysics	Interplay of AAA+ molecular machines, DNA repair enzymes & sliding clamps	3	UCSD
Chemistry	Dynamically tunable ferroelectric surface catalysts	2	Upa
Chemical Sciences	Molecular Simulation of Complex Chemical Systems	2	PNNL
Climate	Simulation of Global Cloudiness	2	ColoradoSU
Fusion Plasma Energy	Gyrokinetic Steady State Transport Simulations	2	Gen Atomics
Fusion Plasma Energy	High Power Electromagnetic Wave Heating in the ITER Burning Plasma	2	ORNL

New algorithm to enable 1+ PFlop/s sustained performance in simulations of disorder effects in high-T_c superconductors

Models,

Methods,

& Implementation

Map to Hardware

Operations

System design

T. A. Maier

P. R. C. Kent

T. C. Schulthess

G. Alvarez

M. S. Summers

E. F. D’Azevedo

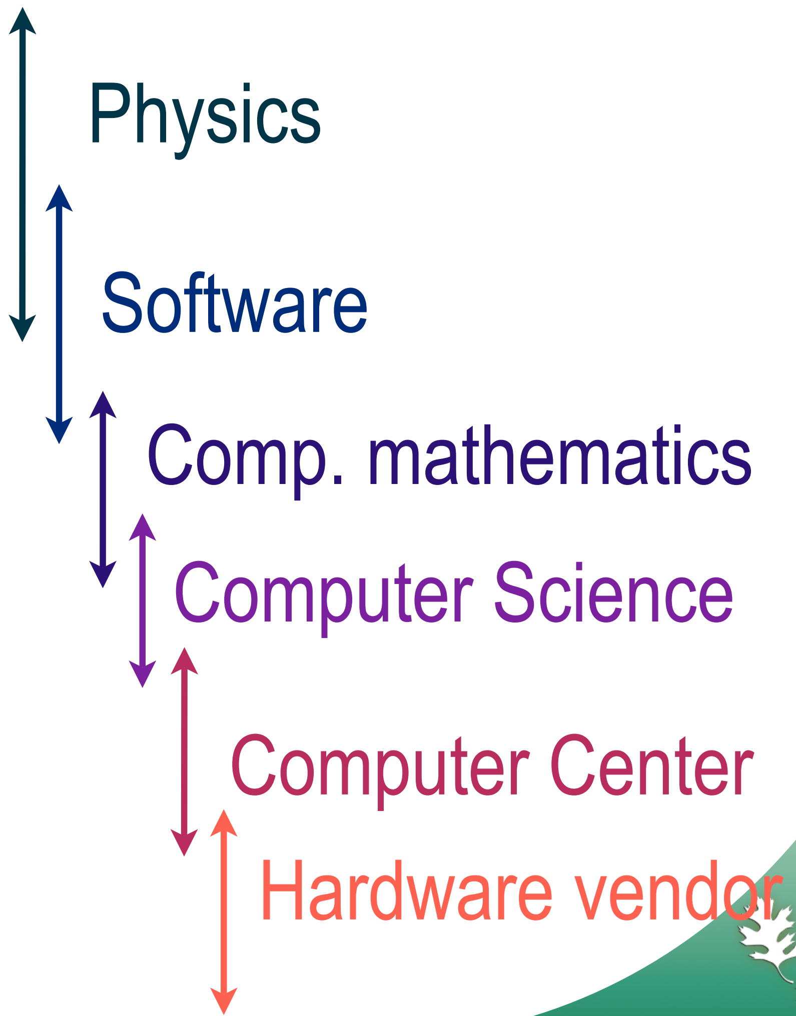
J. S. Meredith

M. Eisenbach

D. E. Maxwell

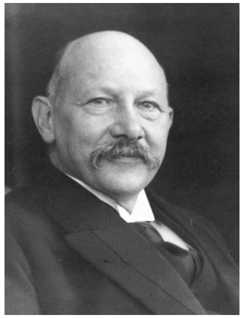
J. M. Larkin

J. Levesque

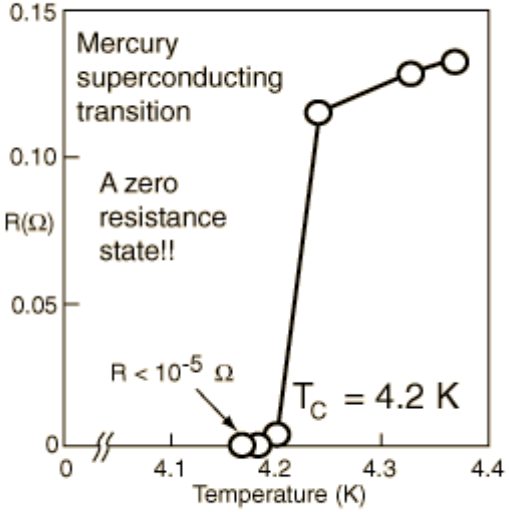


Superconductivity: a state of matter with zero electrical resistivity

Discovery 1911

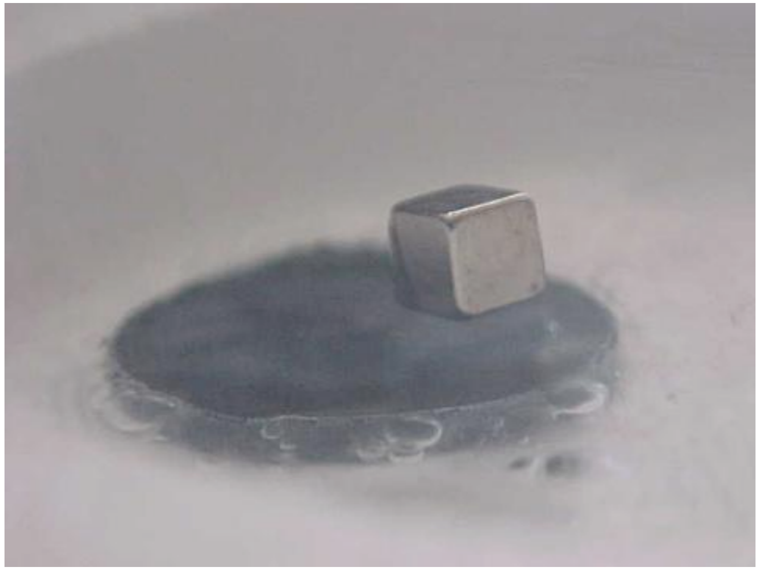


Heike Kamerlingh Onnes (1853-1926)

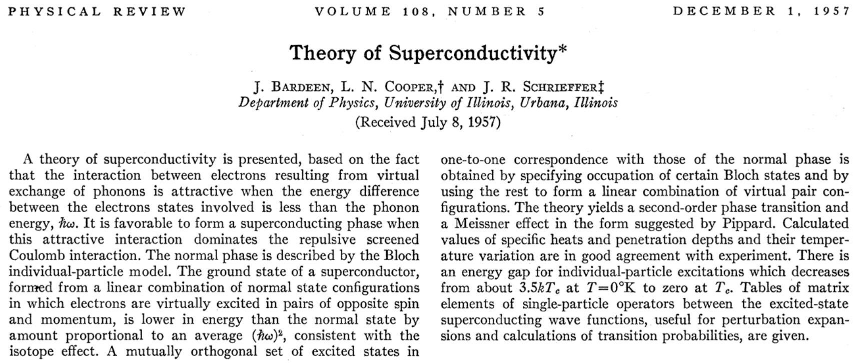


Superconductor repels magnetic field

Meissner and Ochsenfeld, Berlin 1933



Microscopic Theory for Superconductivity 1957



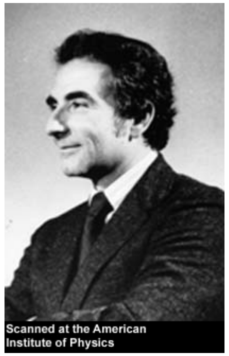
PHYSICAL REVIEW VOLUME 108, NUMBER 5 DECEMBER 1, 1957

Theory of Superconductivity*

J. Bardeen, L. N. Cooper,† and J. R. Schrieffer‡
Department of Physics, University of Illinois, Urbana, Illinois
(Received July 8, 1957)

A theory of superconductivity is presented, based on the fact that the interaction between electrons resulting from virtual exchange of phonons is attractive when the energy difference between the electrons states involved is less than the phonon energy, $\hbar\omega$. It is favorable to form a superconducting phase when this attractive interaction dominates the repulsive screened Coulomb interaction. The normal phase is described by the Bloch individual-particle model. The ground state of a superconductor, formed from a linear combination of normal state configurations in which electrons are virtually excited in pairs of opposite spin and momentum, is lower in energy than the normal state by amount proportional to an average $(\hbar\omega)^2$, consistent with the isotope effect. A mutually orthogonal set of excited states in one-to-one correspondence with those of the normal phase is obtained by specifying occupation of certain Bloch states and by using the rest to form a linear combination of virtual pair configurations. The theory yields a second-order phase transition and a Meissner effect in the form suggested by Pippard. Calculated values of specific heats and penetration depths and their temperature variation are in good agreement with experiment. There is an energy gap for individual-particle excitations which decreases from about $3.5kT_c$ at $T=0°K$ to zero at T_c. Tables of matrix elements of single-particle operators between the excited-state superconducting wave functions, useful for perturbation expansions and calculations of transition probabilities, are given.

Scanned at the American Institute of Physics



Scanned at the American Institute of Physics

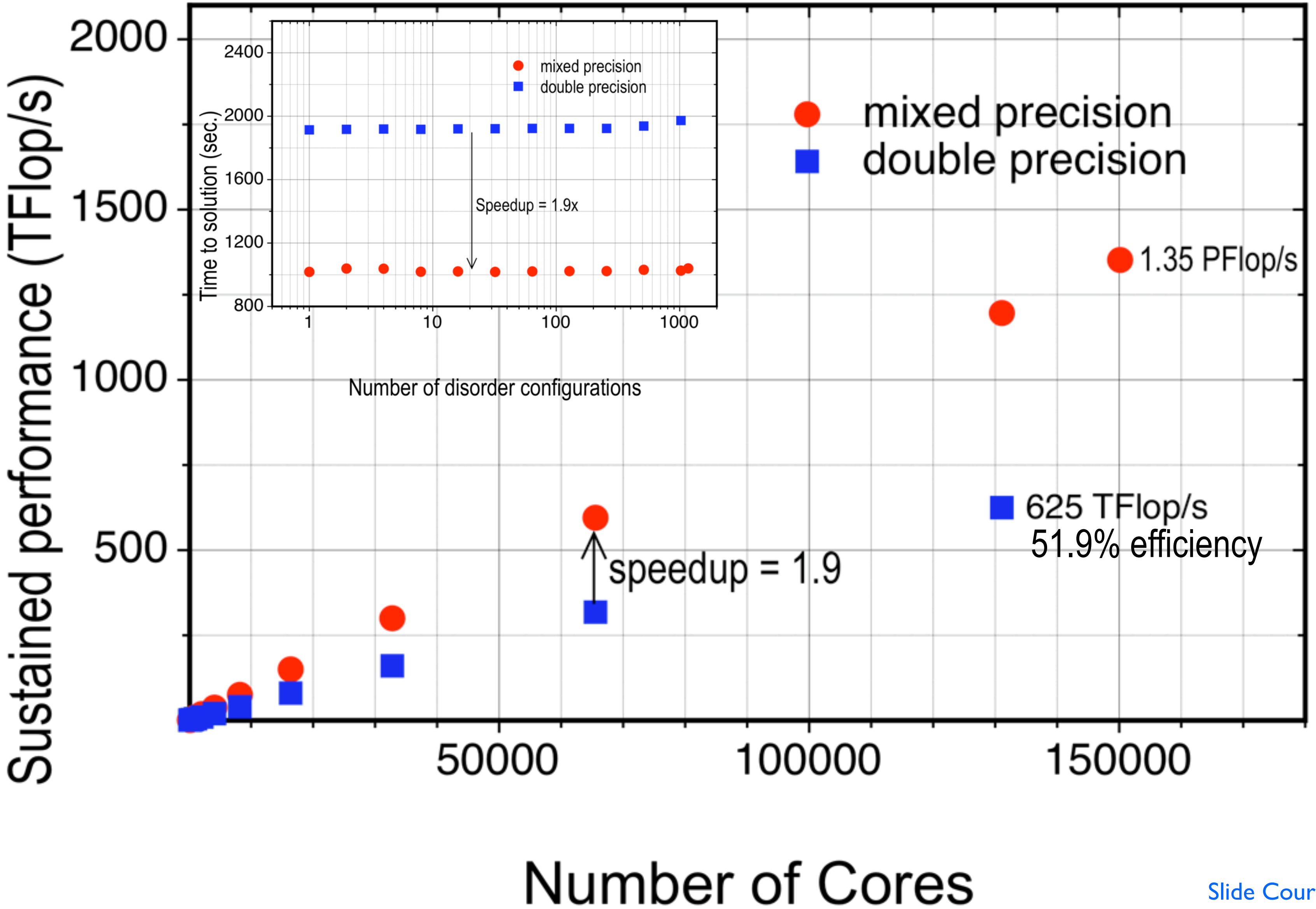
Scanned at the American Institute of Physics

BCS Theory generally accepted in the early 1970s

Courtesy of Thomas Schulthess

Sustained performance of DCA++ on Cray XT5

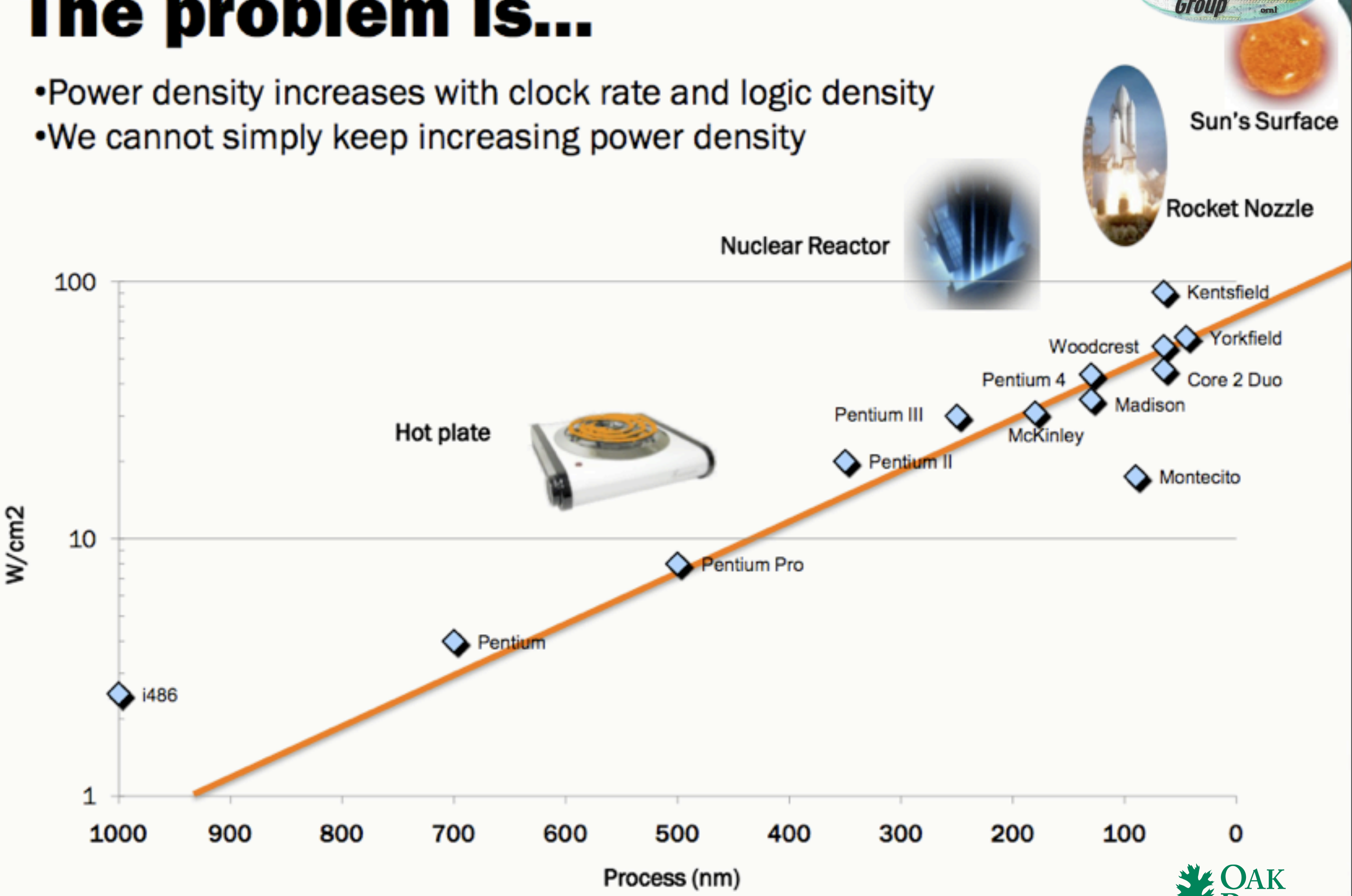
Weak scaling with number disorder configurations, each running on 128 Markov chains on 128 cores (16 nodes) - 16 site cluster and 150 time slides



Slide Courtesy of Thomas Schulthess

The problem is…

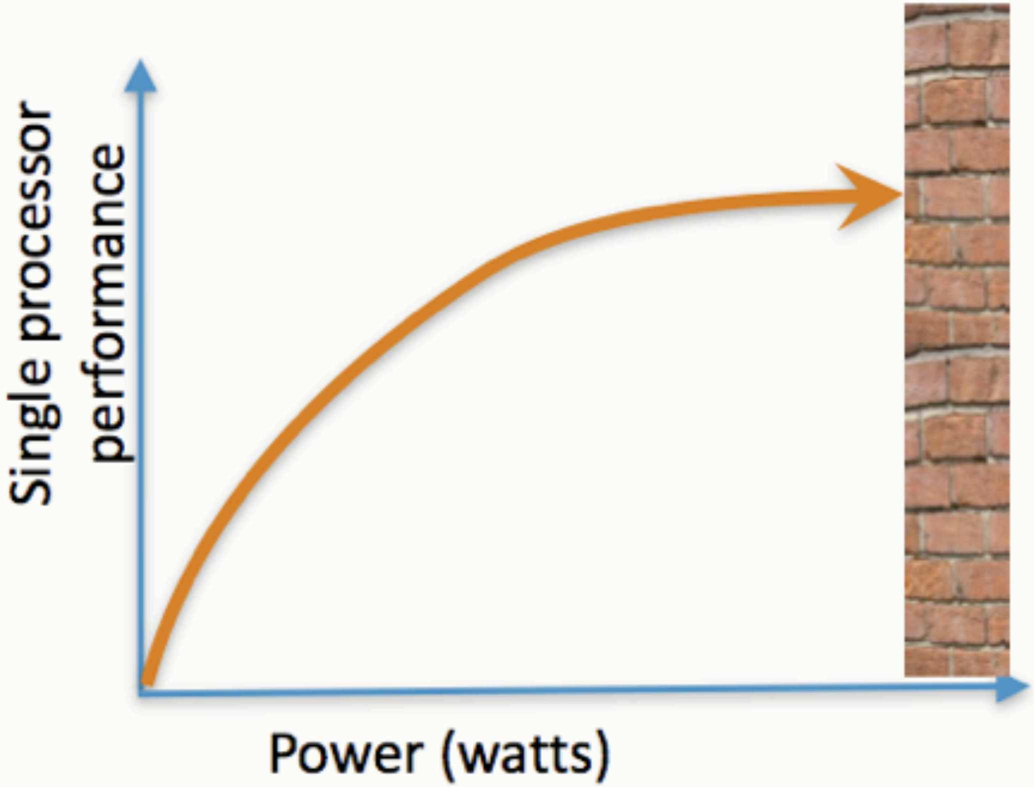
- Power density increases with clock rate and logic density
- We cannot simply keep increasing power density



Adapted from "New Microarchitecture Challenges in the Coming Generations of CMOS Process Technologies," Fred Pollack, Intel Corp. Micro32 conference keynote, 1999

Computing has met a barrier

- In the “Good Old Days” performance doubled every 2 years
  - increased clock rate
  - architectural improvements
- But single threaded performance is increasingly limited by power & cooling



We have hit a “power wall”

Solution: Accelerators

Background: FPGAs & GPUs

Focus: Algorithms => Applications

Goal: Speed Supercomputers with FPGAs

Future Supercomputer Technologies

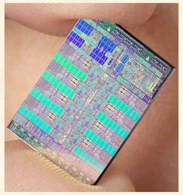
Commodity: 2^n *multi* => *many core*

Special: *El Dorado, Cyclops, PiM*

Accelerators

- **FPGA**: DSP => HPEC => HPC <==

- **Cell**: Sony, Toshiba, IBM



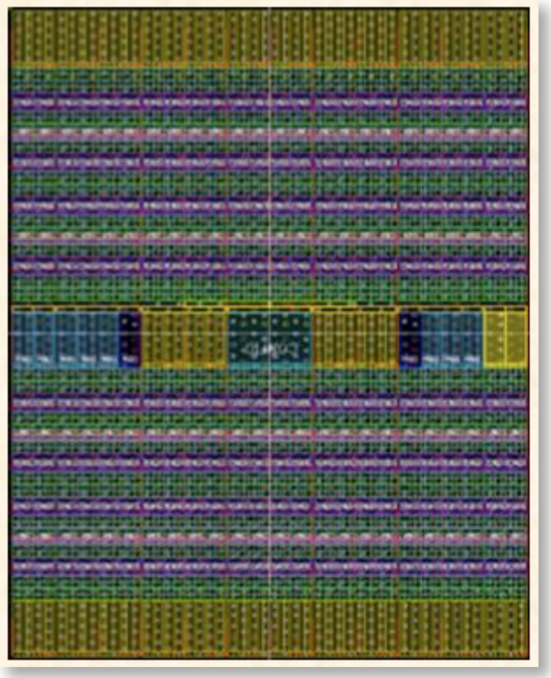
- **GPUs**: => μP

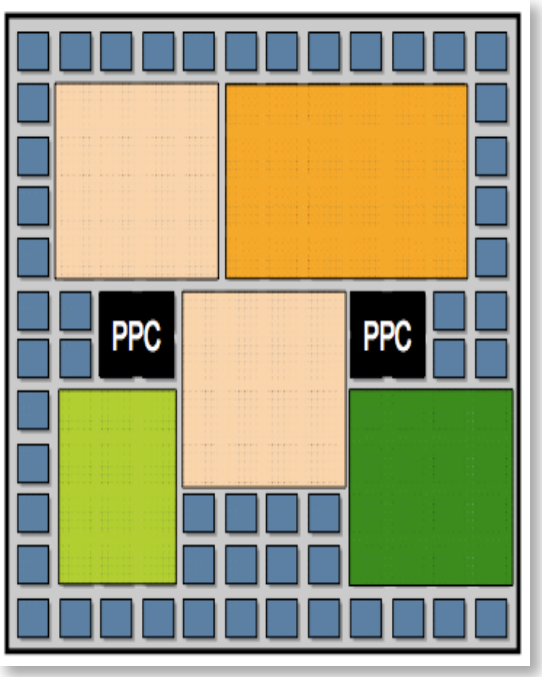
- **Array**: ClearSpeed "niche"

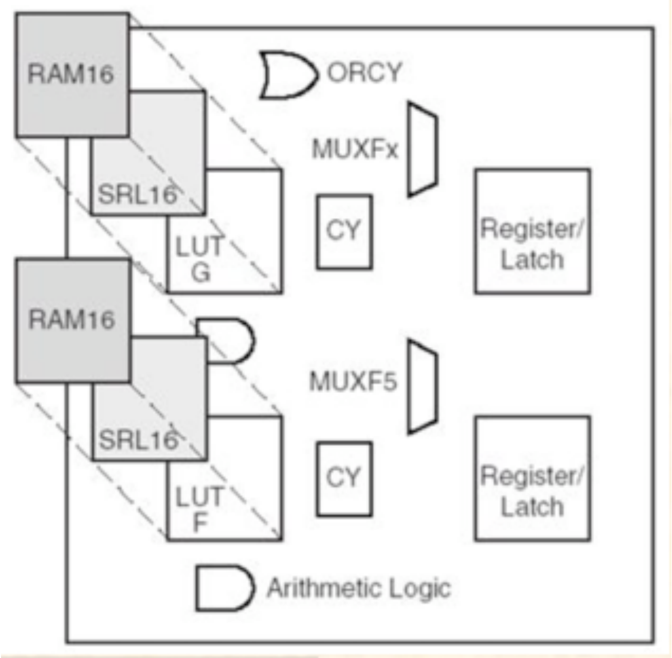
Future Technologies Group ornl

Oak Ridge National Laboratory

What's an FPGA? Your "custom chip"







Xilinx Virtex4 FPGA: 89K slices (miniCPUs)

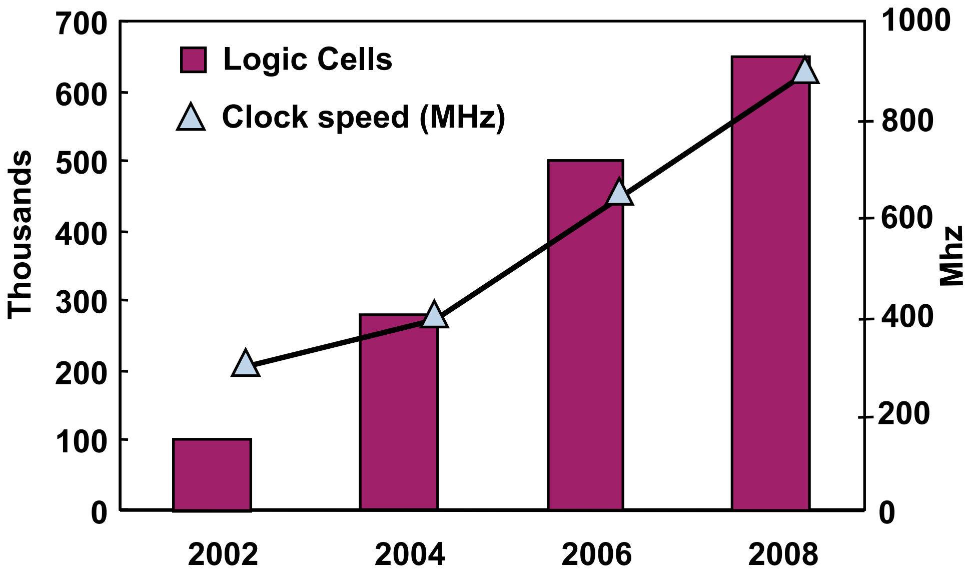
FPGA Logic slice

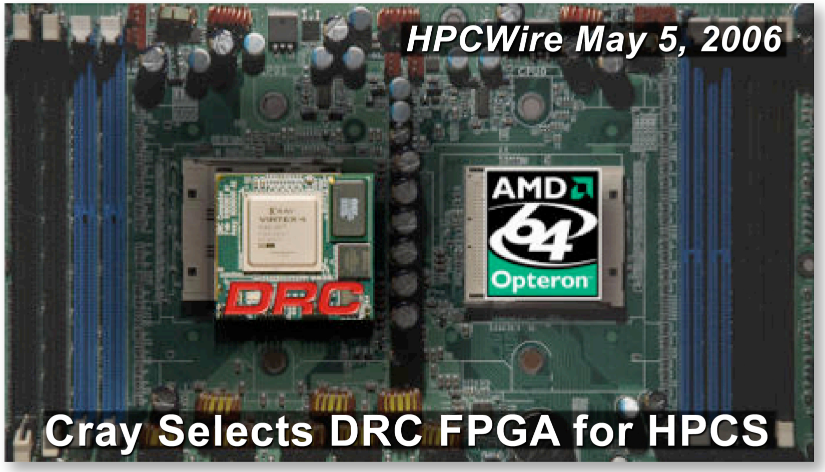
- **Logic array: user-tailored to application**
- **On-chip RAM, multipliers & PowerPCs**
- **Gigabit transceivers/DSP blocks => FastIO/precision**
- **100–1000 operations/clock cycle**

Why FPGAs?

- **Performance**: optimal silicon use (maximize parallel ops/cycle)
- **Rapid growth**: Cells, Speed, I/O
- **Power**: 1/10th CPUs
- **Flexible**: *tailor* to application
- **Advances**: Telecom spinoff

***High clock rate** is a cost, not a benefit; it drives up costs of everything else...*
-- eWeek



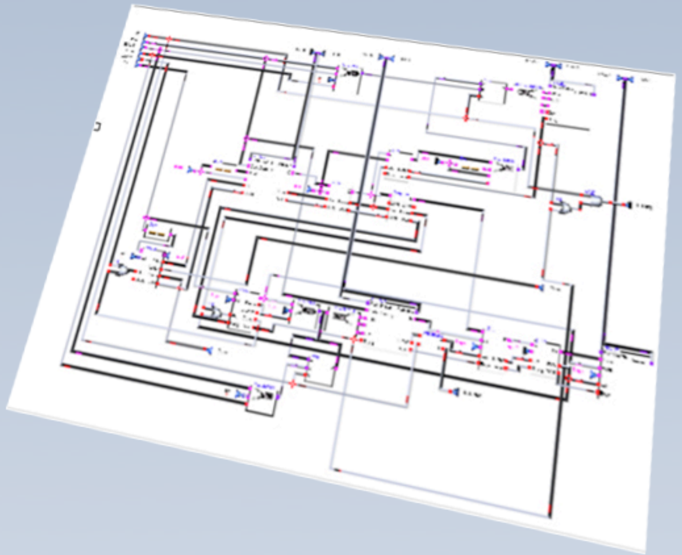


HPCWire May 5, 2006

Cray Selects DRC FPGA for HPCS

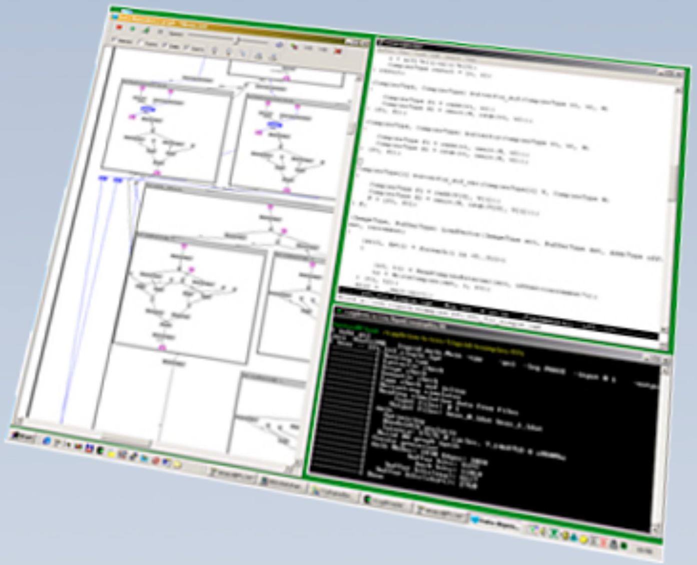
Exploring programming options

Gauss matrix solver



Viva: Graphical Icons—3-dimensional

Compiler, simulator, and debugger



MitrionC: Text/flow—1-dimensional

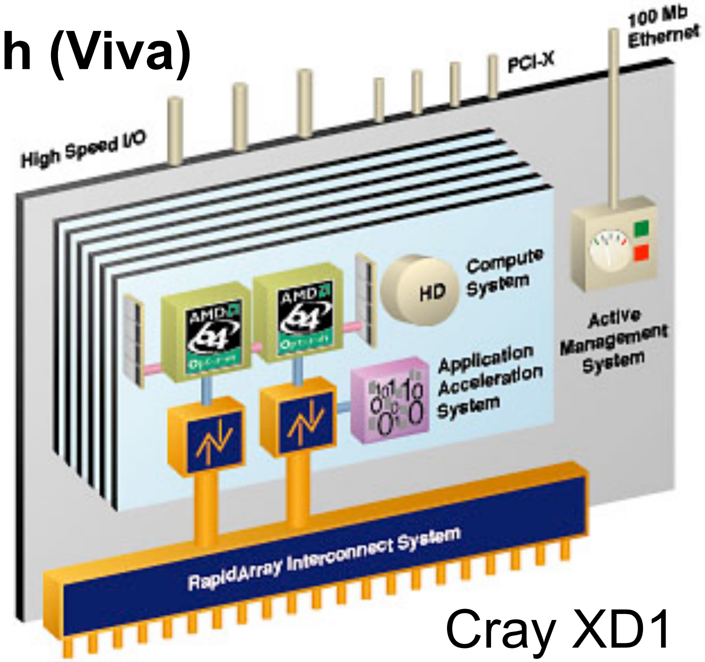
+ Carte/SRC, CHiMPS-VHDL/Xilinx , DSPlogic

Applications

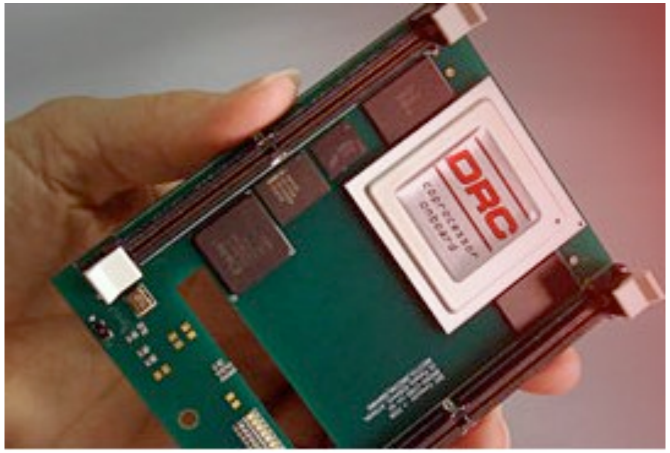
- Genomics
- Matrix Equation Solution
- Molecular Dynamics, Weather/Climate

ORNL FPGA hardware/tools

- SRC-6 (Carte), Digilent (Viva, VHDL), Nallatech (Viva)
- Cray XD1 (MitrionC, VHDL):
  6 FPGAs + 144 Opterons
- SGI RASC-Altix/Virtex4s (MitrionC)
- CHiMPS (Bee2 => Cray XD1 => DRC => XT4)
  (Xilinx early access)

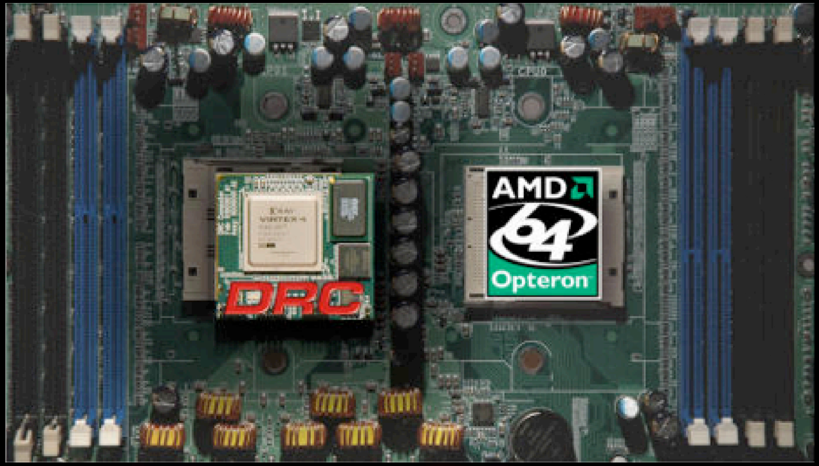


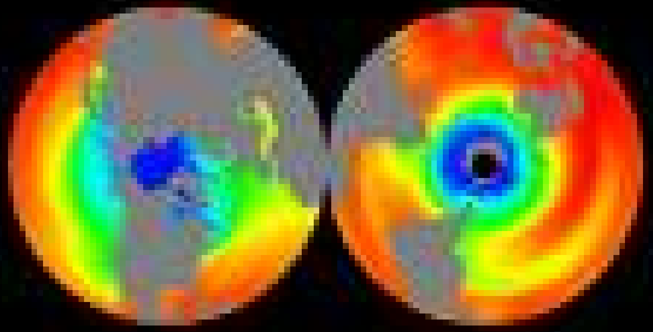
Cray XD1

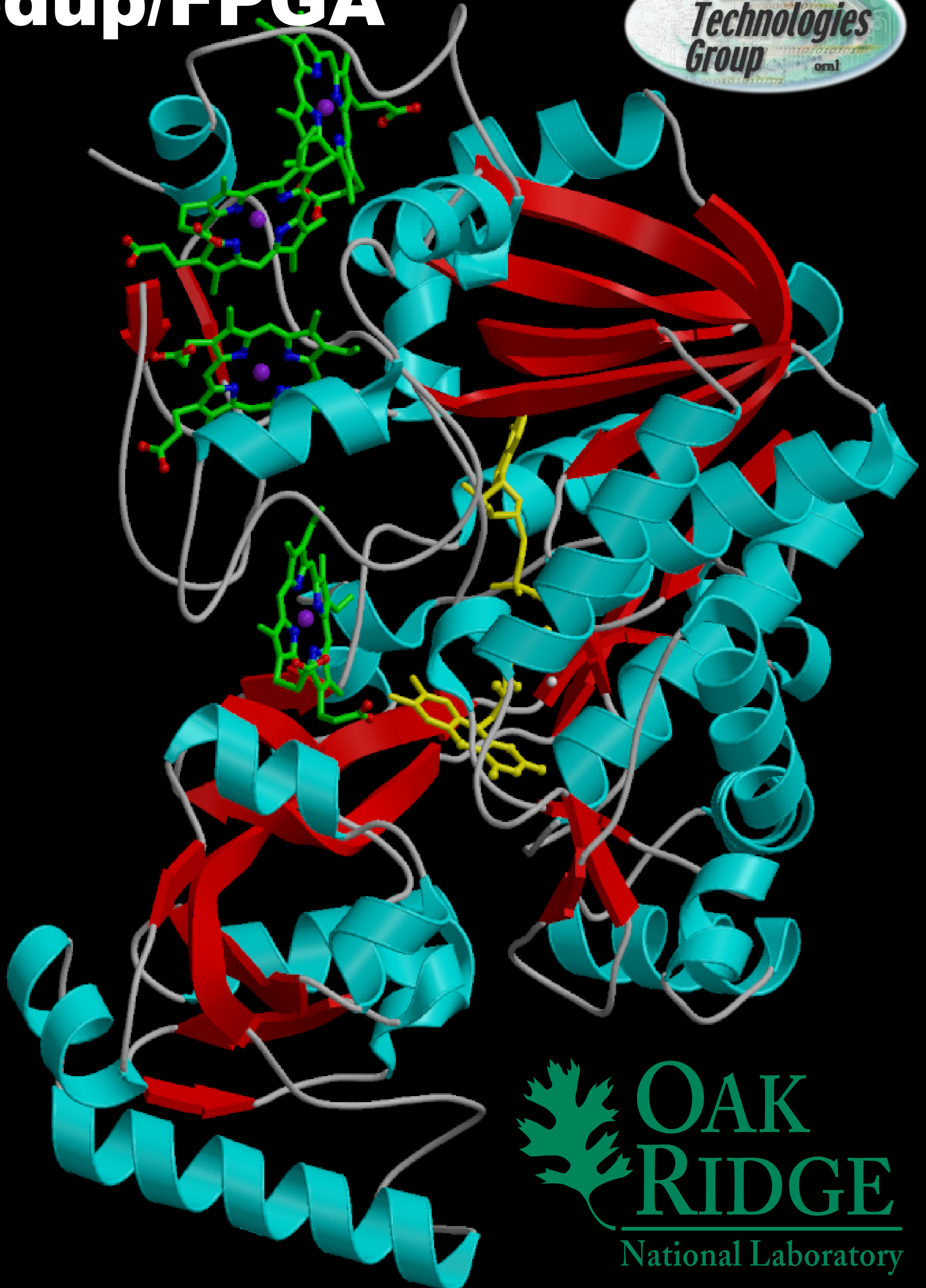


100x Genomics Speedup/FPGA for up to 150 FPGAs

Future Technologies Group ornl







OAK RIDGE
National Laboratory

Openfpga.org Smith-Waterman Benchmark

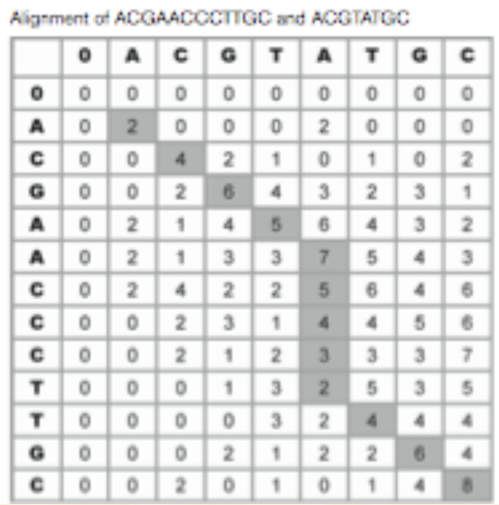
- ***FASTA** (University of Virginia) application*

  http://fasta.bioch.virginia.edu

- **Uses search34 code & Cray SWA core**

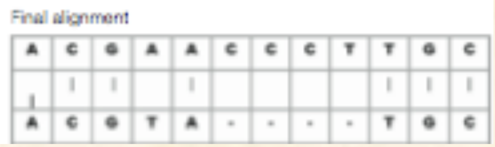
- **Human Genome Data: 4GB compressed**

  3685 searches (MPI on ORNL Cray XD1)



Alignment of ACGAACCCTTGC and ACGTATGC

	0	A	C	G	T	A	T	G	C
0	0	0	0	0	0	0	0	0	0
A	0	2	0	0	0	2	0	0	0
C	0	0	4	2	1	0	1	0	2
G	0	0	2	6	4	3	2	3	1
A	0	2	1	4	5	6	4	3	2
A	0	2	1	3	3	7	5	4	3
C	0	2	4	2	2	5	6	4	6
C	0	0	2	3	1	4	4	5	6
C	0	0	2	1	2	3	3	3	7
T	0	0	0	1	3	2	5	3	5
T	0	0	0	0	3	2	4	4	4
G	0	0	0	2	1	2	2	6	4
C	0	0	2	0	1	0	1	4	8



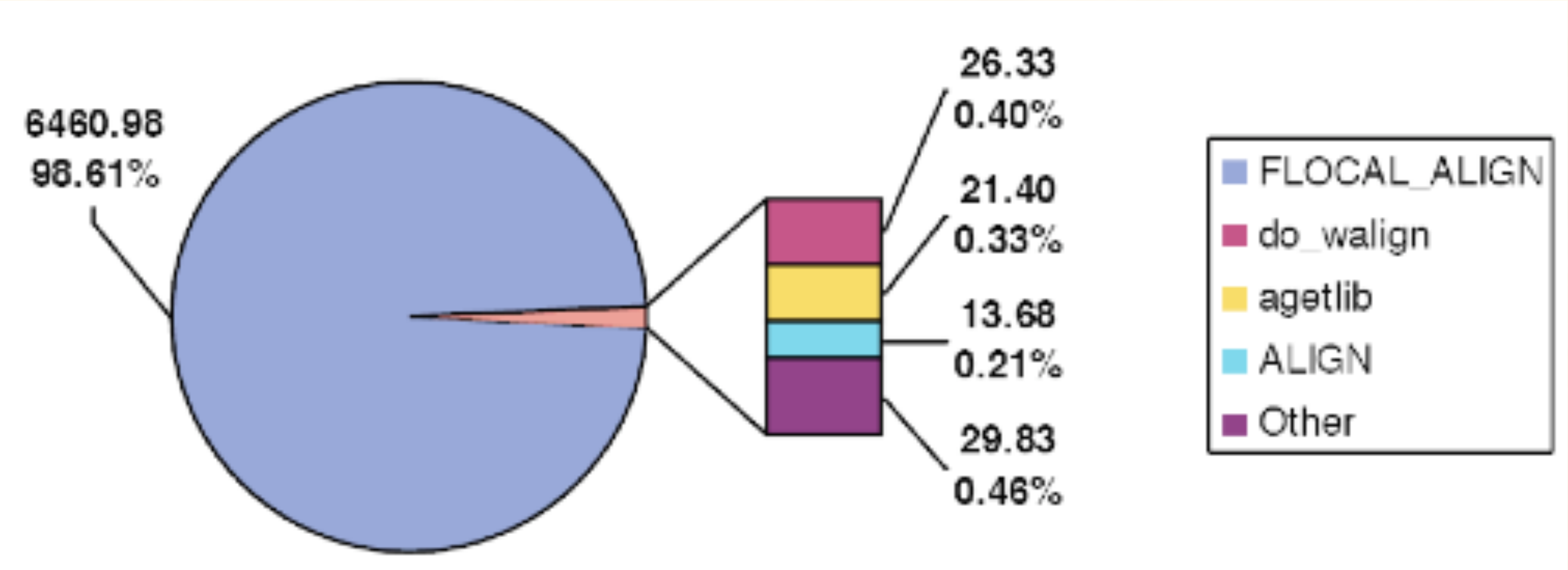
Final alignment

A	C	G	A	A	C	C	C	T	T	G	C
|	|	|		|					|	|	|
A	C	G	T	A	-	-	-	-	T	G	C

Future Technologies Group ornl

OAK RIDGE National Laboratory

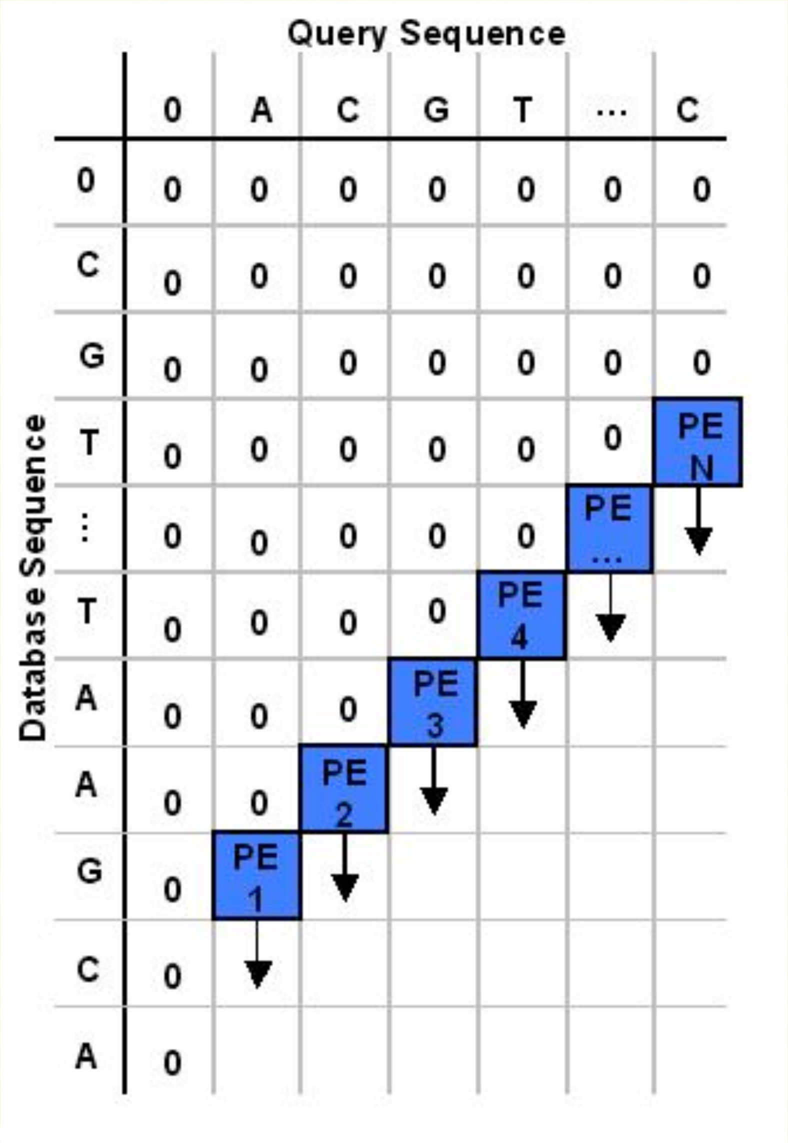
Search34 Computation Profile



98.61% is FLOCAL_ALIGN **=> VHDL kernel**

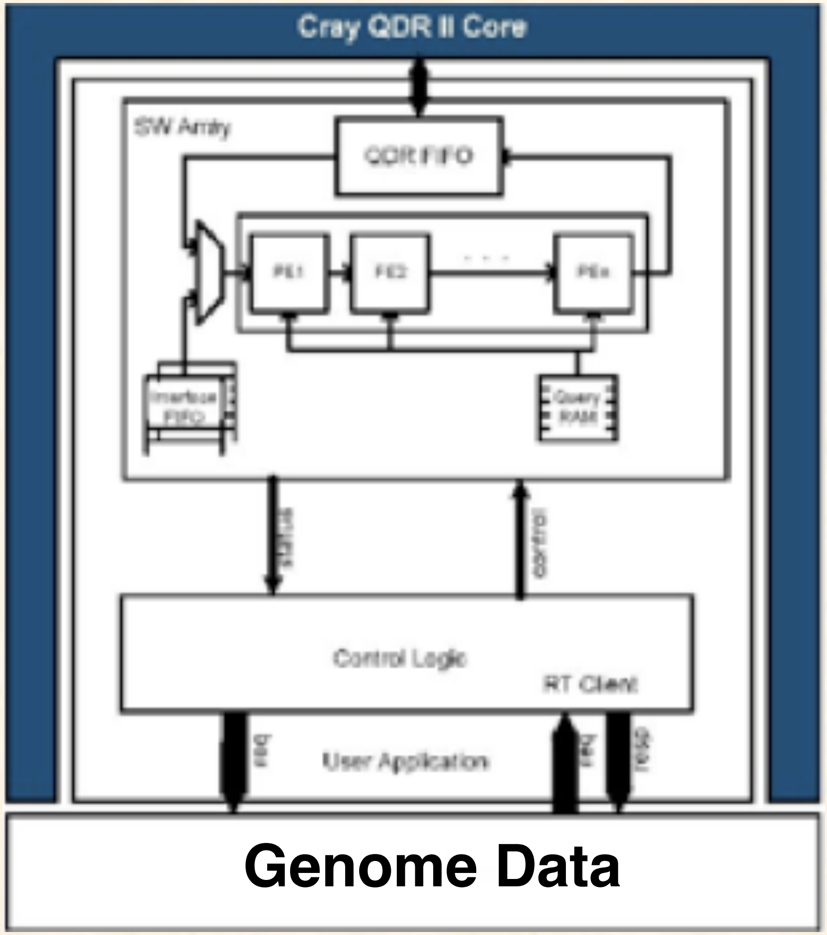
Smith-Waterman

Parallel Score Calculation



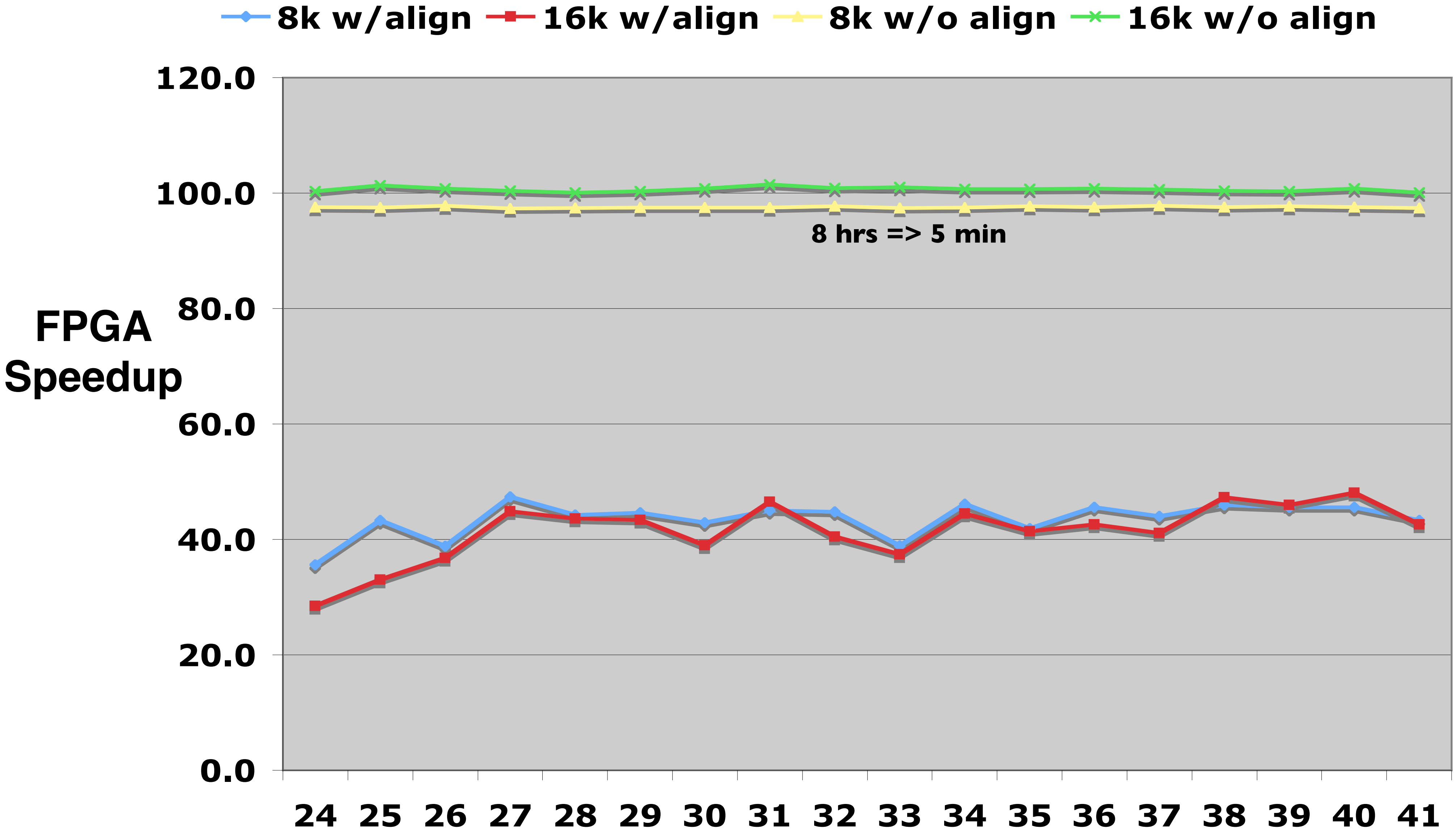
	0	A	C	G	T	…	C
0	0	0	0	0	0	0	0
C	0	0	0	0	0	0	0
G	0	0	0	0	0	0	0
T	0	0	0	0	0	0	PE N
⋮	0	0	0	0	0	PE …	
T	0	0	0	0	PE 4		
A	0	0	0	PE 3			
A	0	0	PE 2				
G	0	PE 1					
C	0						
A	0						

Overall Algorithm



100x* DNA Sequence Speedup

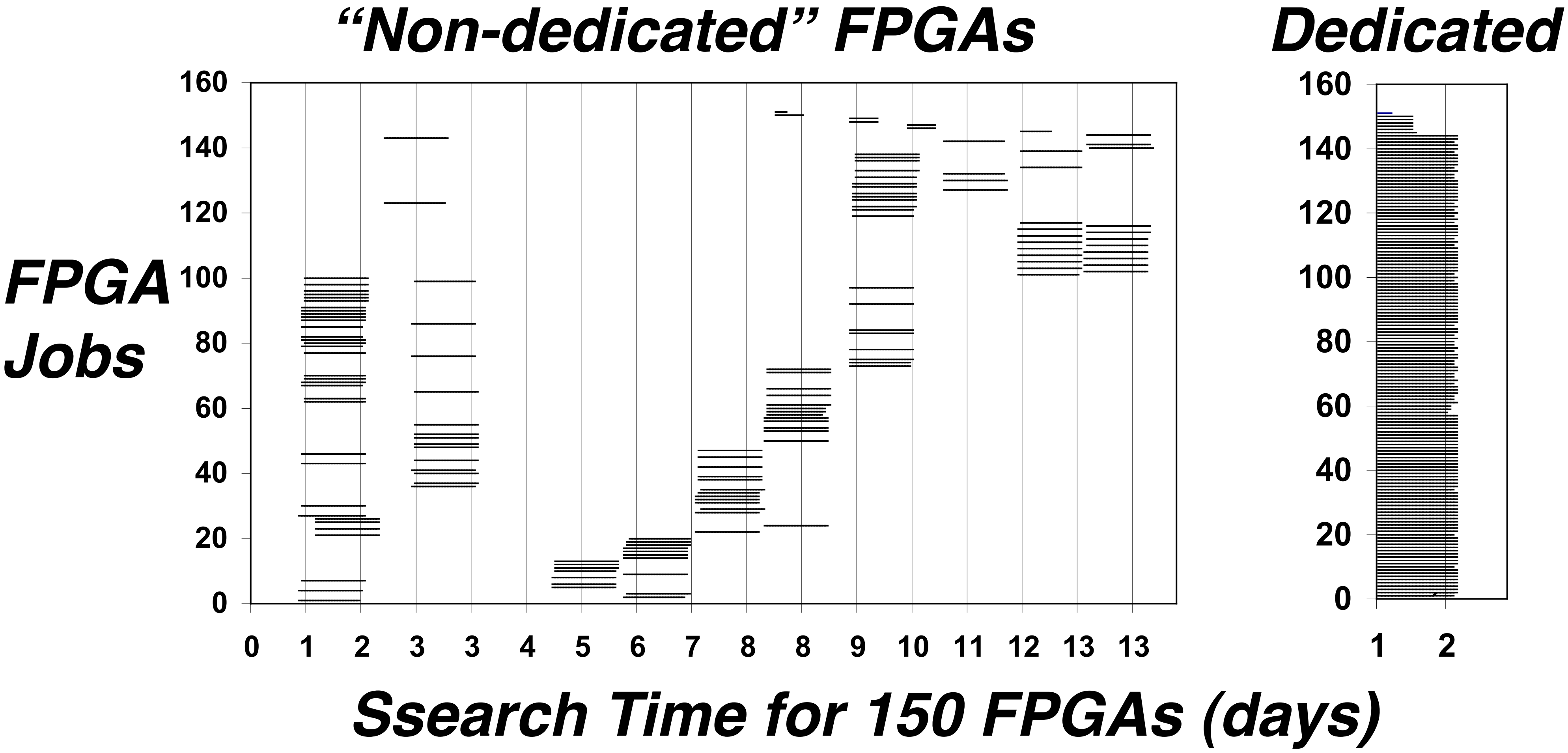
Bacillus anthracis Human DNA comparison



*Virtex-4 FPGA vs 2.2 GHz Opteron on Cray XD1

DNA Sequencing* Time on 150 FPGAs

*Human-Mouse DNA Compare (FASTA)



DNA Sequencing: Speed* on 150 FPGAs

*State-of-the-art: **G**iga **C**ell **U**pdates **P**er **S**econd (**GCUPS**)

- DNA Characters: **Human = 155 million, Mouse = 165 million**

Total Compares = 155M x 165M x 106^2 x 2
= 51×10^{15} Cell Updates

- Sequential FPGA ==> 138 days (11,923,200 secs) **==> 4.3 TCUPS**
  (**51×10^{15}/11,923,200 Tera CUPS)**
- Parallel (actual) ==> 12.9 days (1,114,560 secs) **==> 46 TCUPS**
- Parallel (dedicated) ==> 1 day (86,400 secs) **==> 605 TCUPS**

Speedup on 150 FPGAs*

1 Opteron ==> 20 years (240 mos)

1 FPGA ==> 5 months

150 Opterons ==> 6 weeks

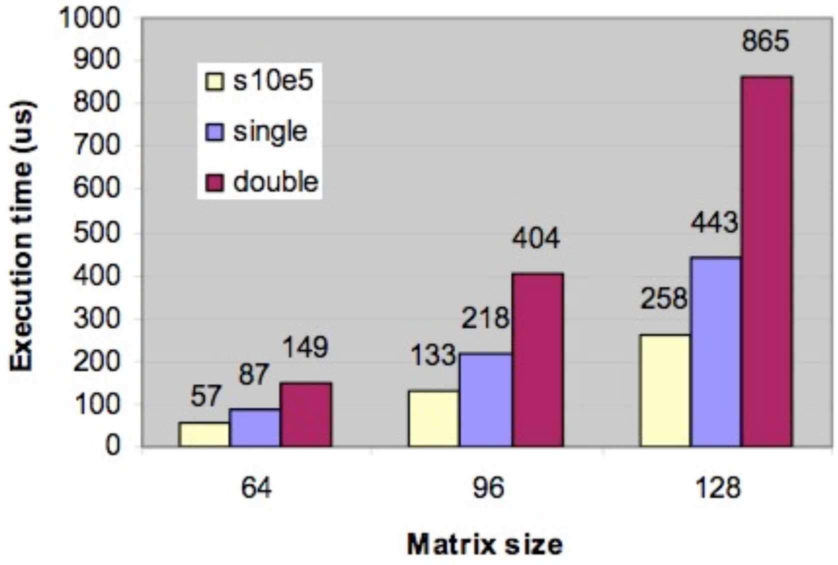
150 FPGAs ==> 1 day ==> 49X speedup (VirtexII)

==> 7,350X faster than 1 Opteron (VirtexIIs)

==> 14,700X faster than 1 Opteron (Virtex4s)

*Compared to one 2.2 GHz Opteron

37x* LU Decomposition FPGA Speedup
10x for Matrix Equation Solver



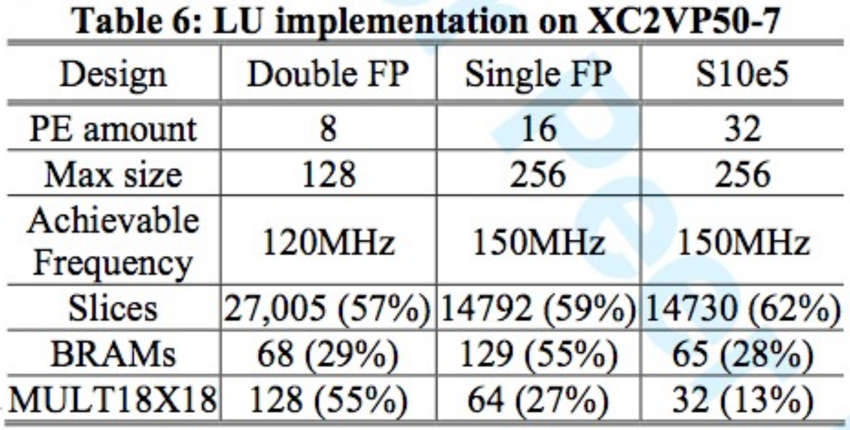


Table 6: LU implementation on XC2VP50-7

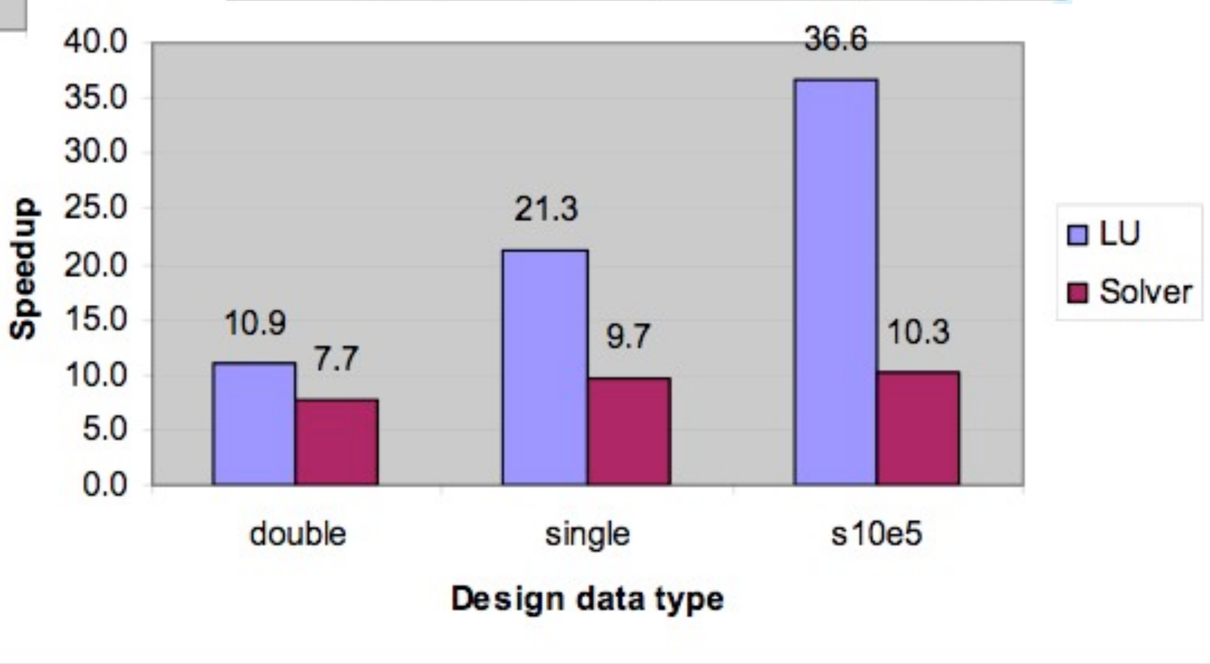
Design	Double FP	Single FP	S10e5
PE amount	8	16	32
Max size	128	256	256
Achievable Frequency	120MHz	150MHz	150MHz
Slices	27,005 (57%)	14792 (59%)	14730 (62%)
BRAMs	68 (29%)	129 (55%)	65 (28%)
MULT18X18	128 (55%)	64 (27%)	32 (13%)

Benefits:

High performance **of LP arithmetic**

High precision **accuracy**

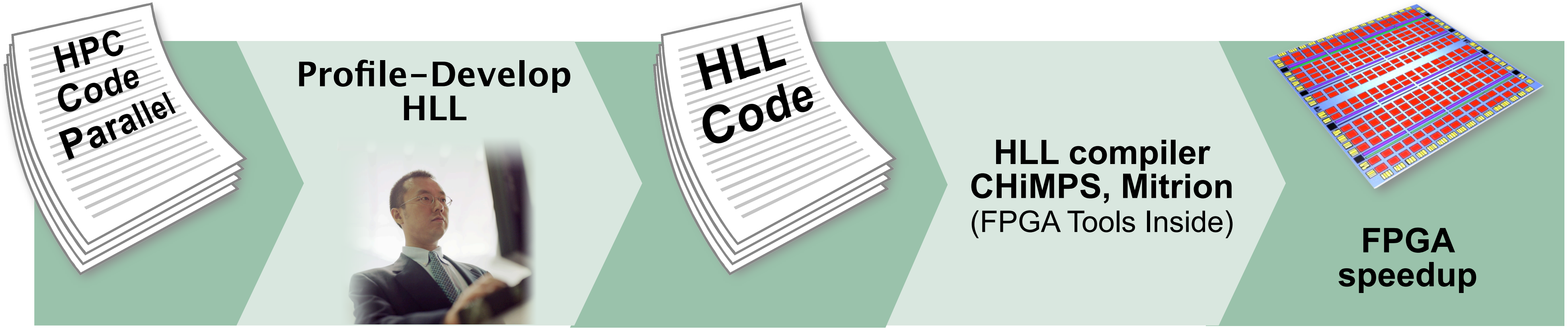
Speedup increases **with matrix size**
(LU dominates calculations)



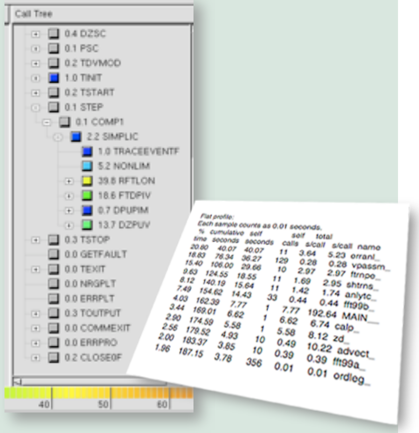
First mixed-precision LU & solver for FPGAs

*Virtex-II vs 2.2 GHz Opteron

Ported Weather-Climate code Spectral Transform Shallow Water Model (STSWM) **to FPGAs**



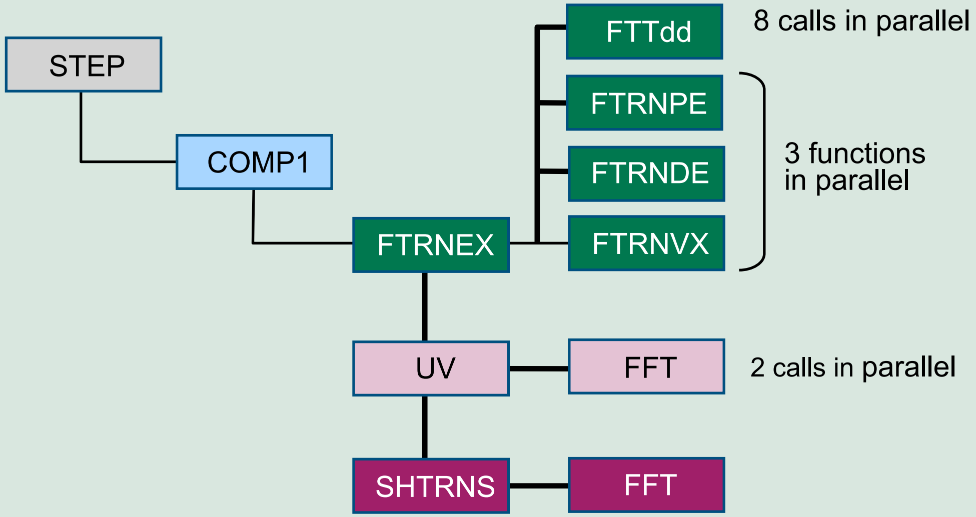
Profile



```
Call Tree
 + 0.4 DZSC
 + 0.1 PSC
 + 0.2 TDVMOD
 + 1.0 TINIT
 + 0.2 TSTART
 - 0.1 STEP
   - 0.1 COMP1
     - 2.2 SIMPLIC
         1.0 TRACEEVENTF
         5.2 NONLIM
       + 39.8 RFTLON
       + 18.6 FTDPIV
       + 0.7 DPUPIM
       + 13.7 DZPUV
 + 0.3 TSTOP
   0.0 GETFAULT
 + 0.0 TEXIT
   0.0 NRGPLT
   0.0 ERRPLT
 + 0.3 TOUTPUT
 + 0.0 COMMEXIT
 + 0.0 ERRPRO
 + 0.2 CLOSE0F
40        50        60
```

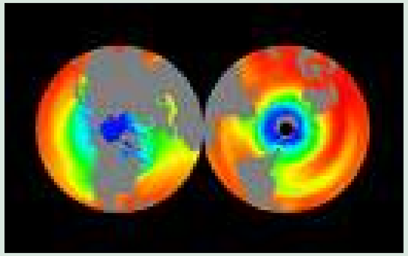
```
Flat profile:
Each sample counts as 0.01 seconds.
  %   cumulative   self              self     total
 time   seconds   seconds    calls  s/call   s/call  name
 20.80     40.07    40.07       11    3.64     5.23  orranl_
 18.83     76.34    36.27      129    0.28     0.28  vpassm_
 15.40    106.00    29.66       10    2.97     2.97  ftrnpe_
  9.63    124.55    18.55       11    1.69     2.95  shtrns_
  8.12    140.19    15.64       11    1.42     1.74  anlytc_
  7.49    154.62    14.43       33    0.44     0.44  fft99b_
  4.03    162.39     7.77        1    7.77   192.64  MAIN__
  3.44    169.01     6.62        1    6.62     6.74  calp_
  2.90    174.59     5.58        1    5.58     8.12  zd_
  2.56    179.52     4.93       10    0.49    10.22  advect_
  2.00    183.37     3.85       10    0.39     0.39  fft99a_
  1.96    187.15     3.78      356    0.01     0.01  ordlog_
```

Find parallelism: 80% FFTs



Goal

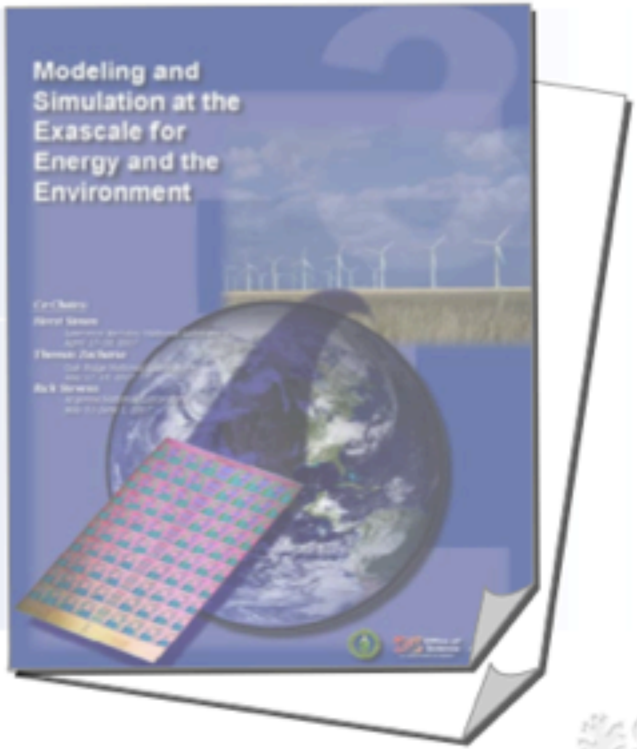
More GF/$ GF/Watt



Model 5-10X faster

Exascale computing and the resiliency challenge

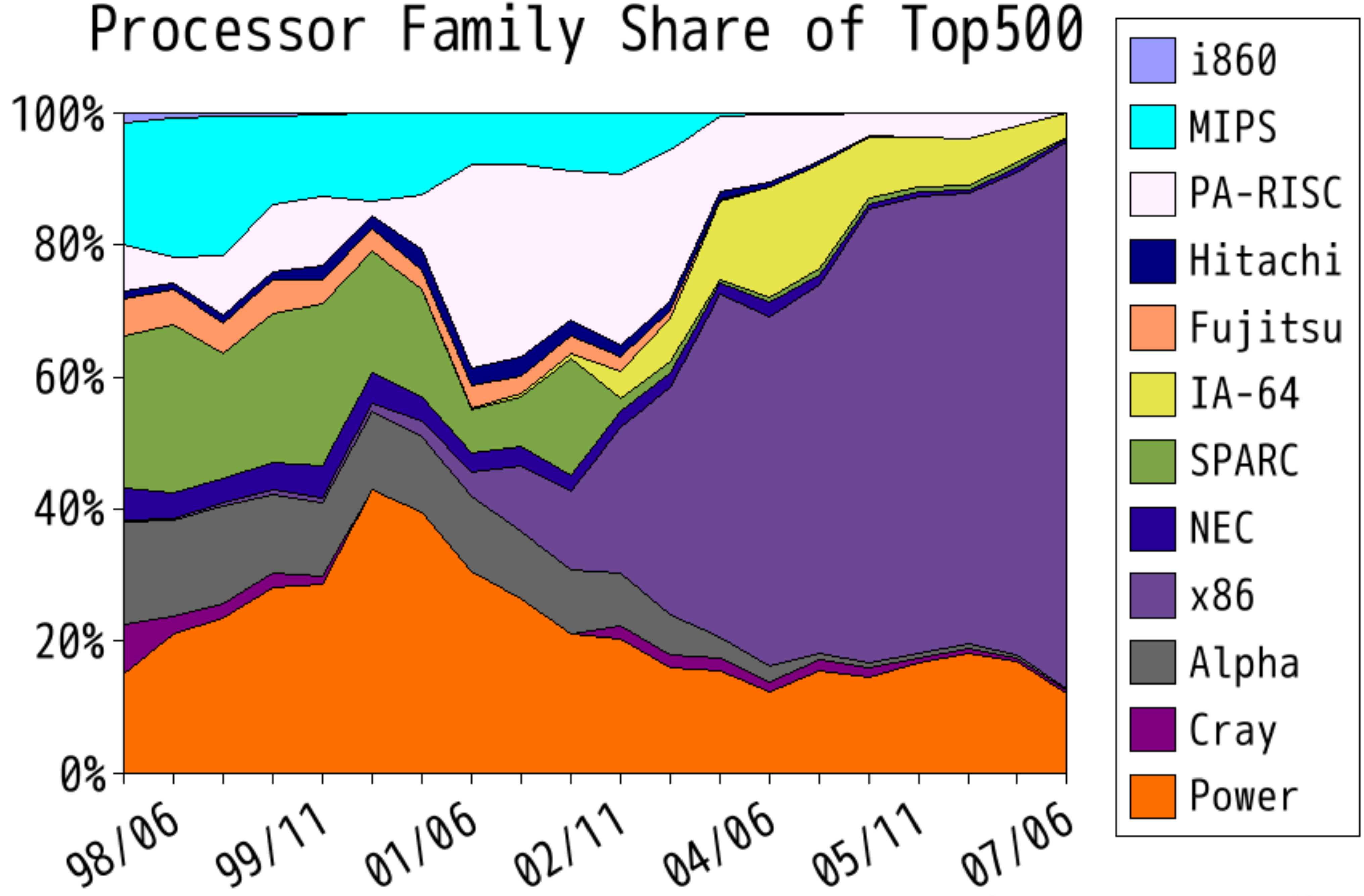
Climate	Improve our understanding of complex biogeochemical cycles that underpin global ecosystem functions and control the sustainability of life on Earth
Energy	Develop and optimize new pathways for renewable energy production and development of long-term, secure nuclear energy sources, optimize energy efficiency, understand "water."
Biology	Enhance our understanding of the roles and functions of microbial life on Earth, and adapt these capabilities for human use. Understand "water."
Socioeconomics	Develop integrated modeling environments for coupling the wealth of observational data and complex models to economic, energy, and resource models



Modeling and
Simulation at the
Exascale for
Energy and the
Environment

Slide courtesy of Thomas Zacharia

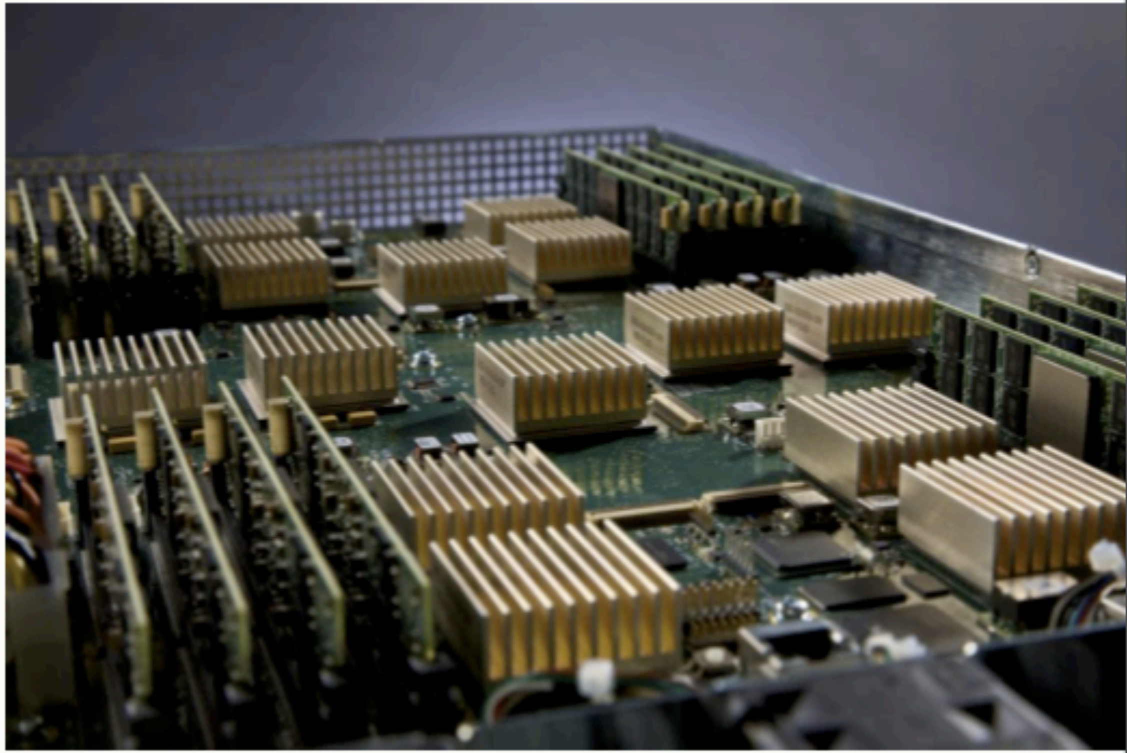
Processor Family Share of Top500



Source: http://www.top500.org

Performance of Application Specific Hardware

- Increased memory bandwidth and processing capability
- Dynamically reloadable with application specific functions ("personalities")

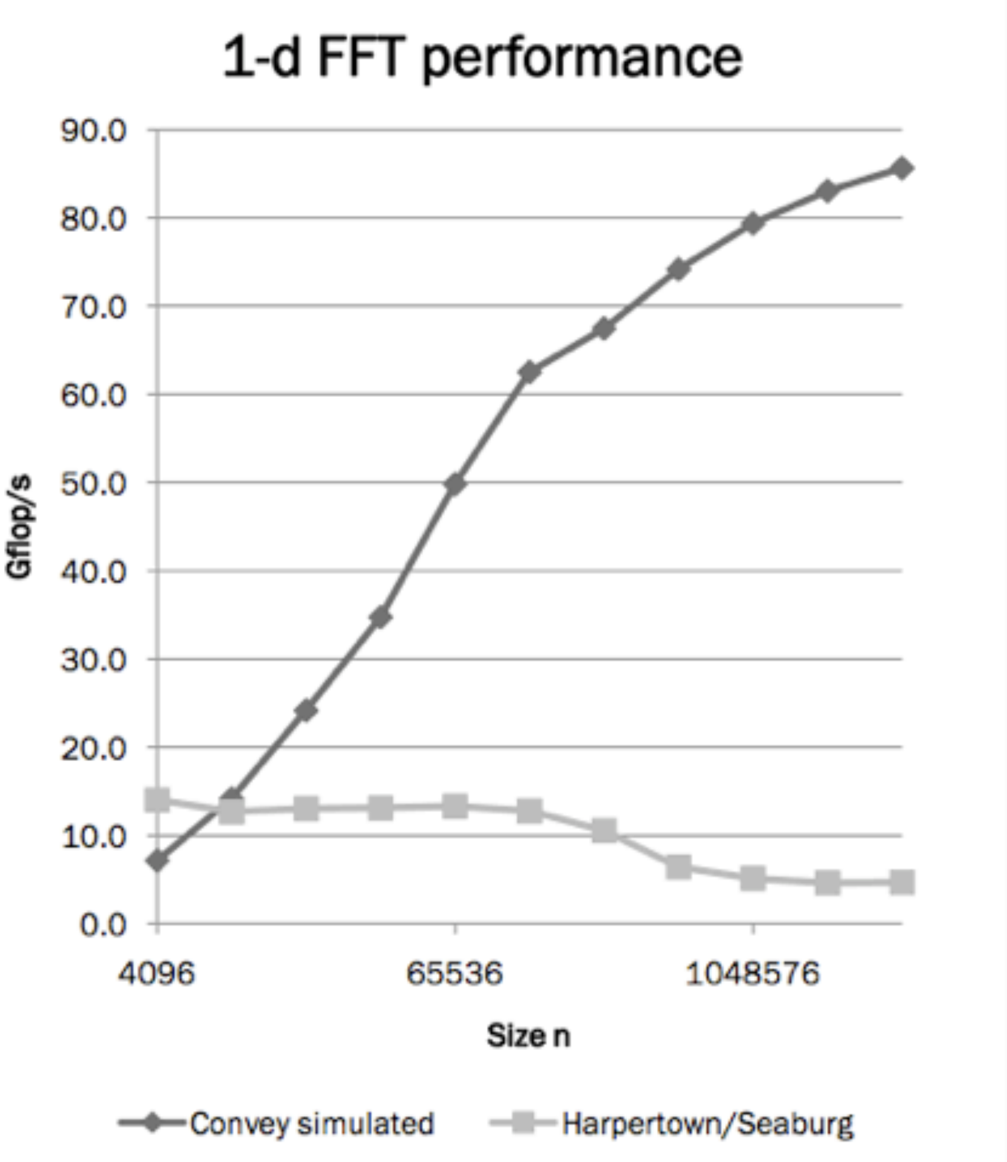


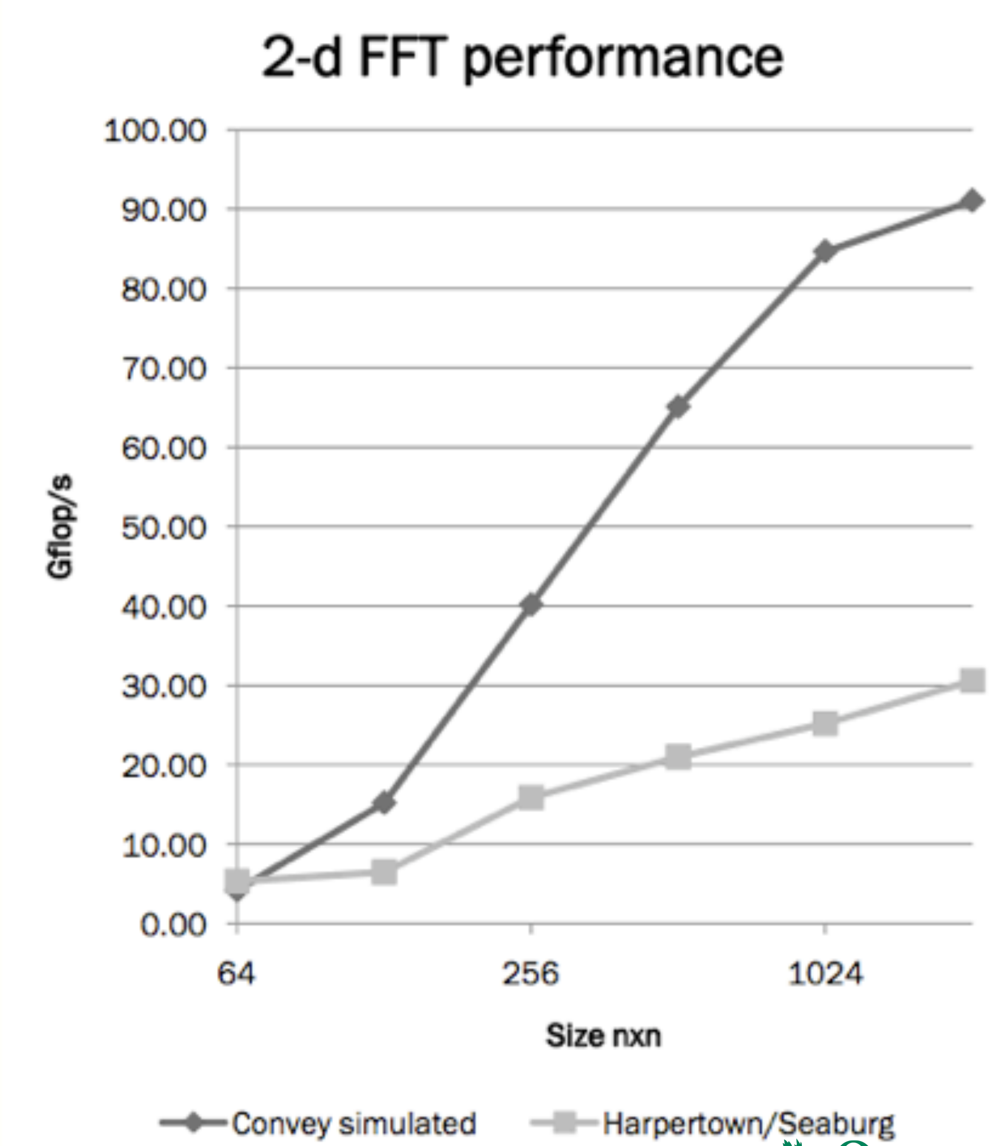
The performance of one rack of Convey Hybrid-Core Computers

the performance of 6 or more racks of commodity servers

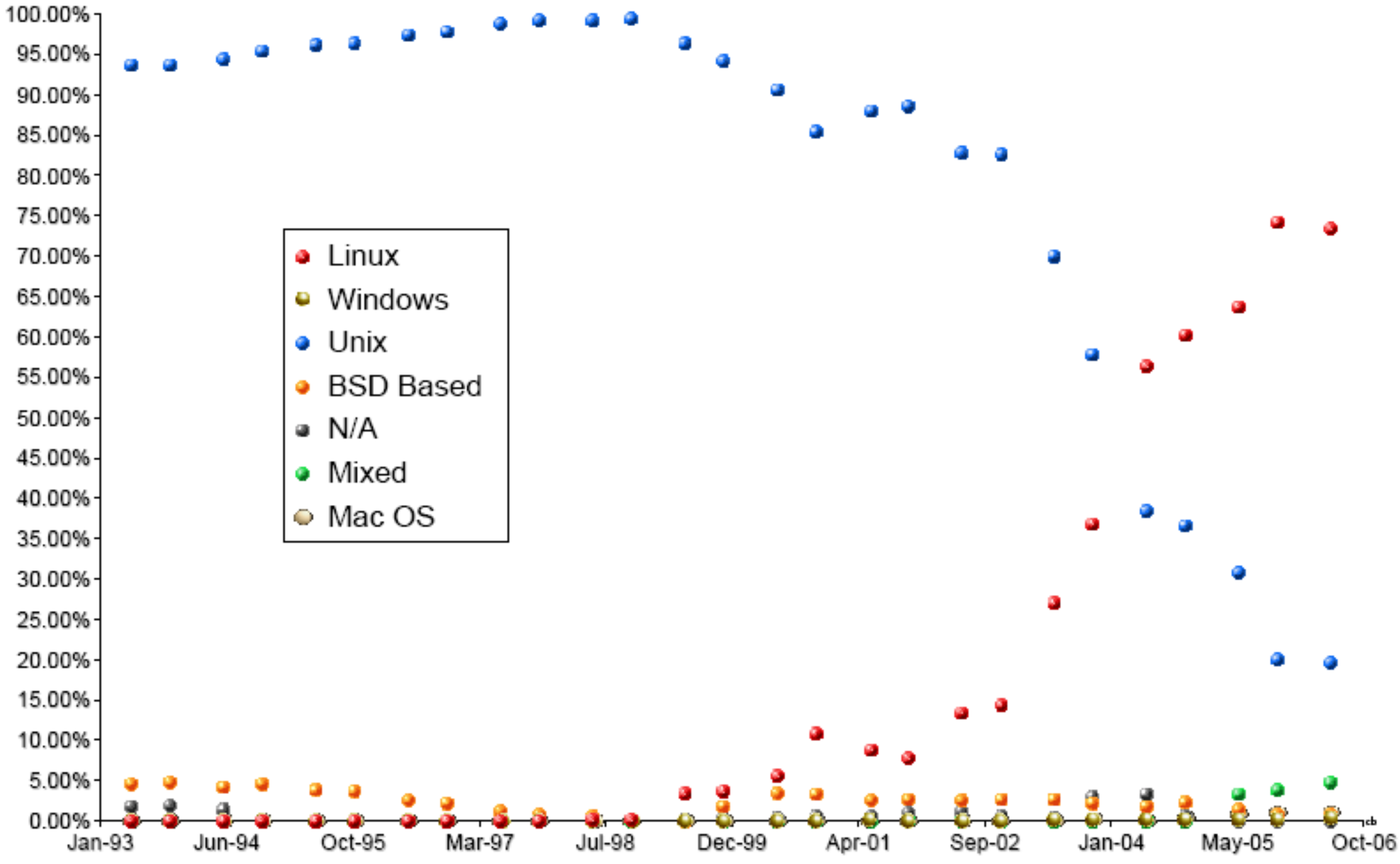
Provides higher absolute performance and more performance per dollar, watt, and unit of floor space

FFT Performance with the SPvector personality





Operating Systems Used On Top500 Supercomputers



Source: www.top500.org

Summary

- ORNL HPC & FPGA research:

  - ORNL Tops in Supercomputing for Science

    (3 PetaFLOP supercomputers - planning ExaFLOP)

  - GPUs & FPGAs growth in HPC

  - Partners: *Cray, Xilinx, UT, NRL, NVidia, SGI, Convey*

  - Speedup: 10X Eqn Soln, 100X/FPGA DNA Sequencing

  - Scalable: to 150 FPGAs (Genomics)

- ORNL hiring

Acknowledgment: This U.S Government work (public domain) was supported by the Office of Science, U. S. Department of Energy Contract DE-AC05-00OR22725. The authors also thank the US Naval Research Lab. for access to the 150 FPGA Cray XD.

Contact

Olaf O. Storaasli
Future Technologies Group
Google **Olaf ORNL**

THANK YOU