LA-SiGMA

Louisiana Alliance for Simulation-Guided Materials Applications

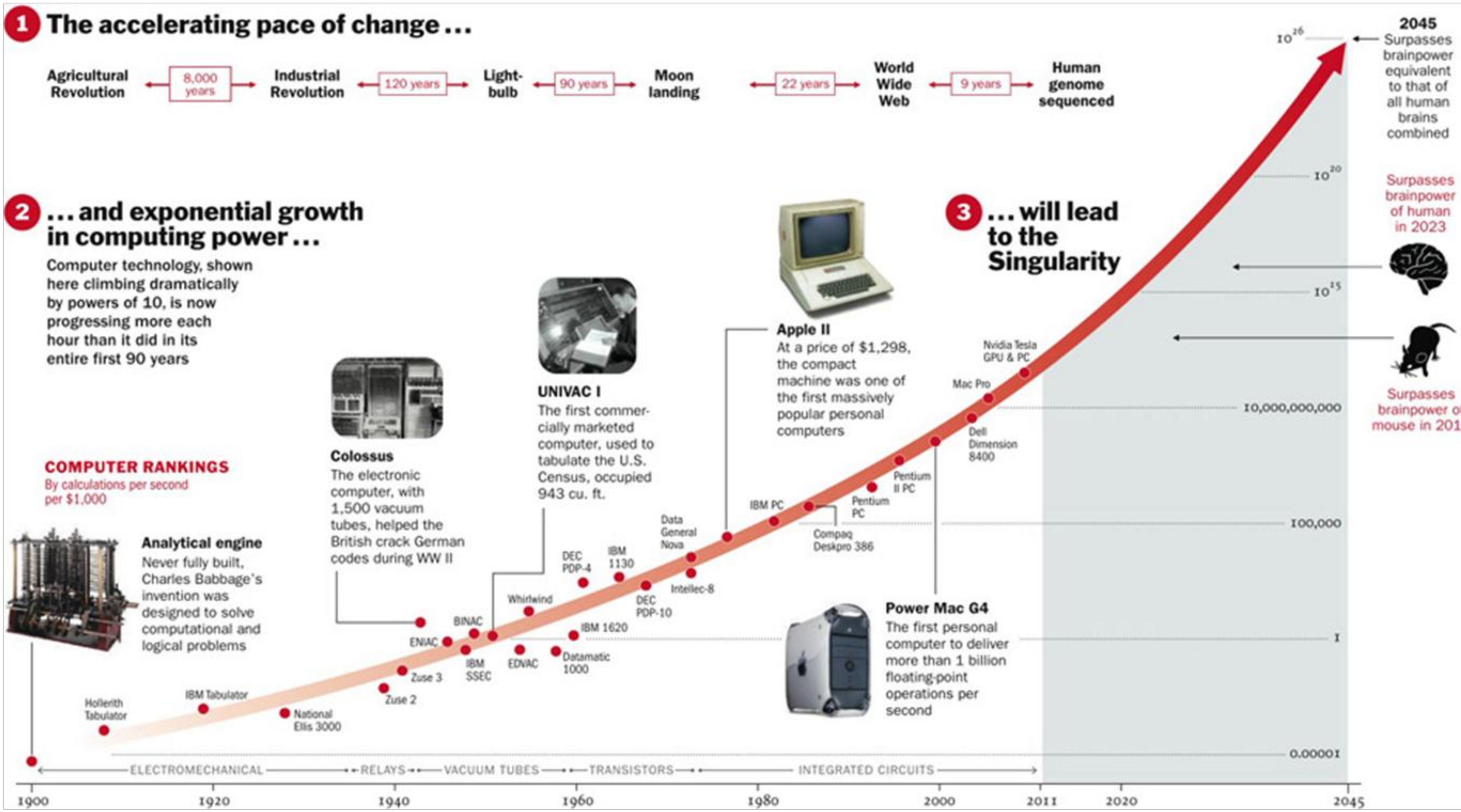
Computational Teams Cybertools/Cyberinfrastructure (CTCI)

"The glue"

All-Hands Meeting, Baton Rouge: August 5, 2011

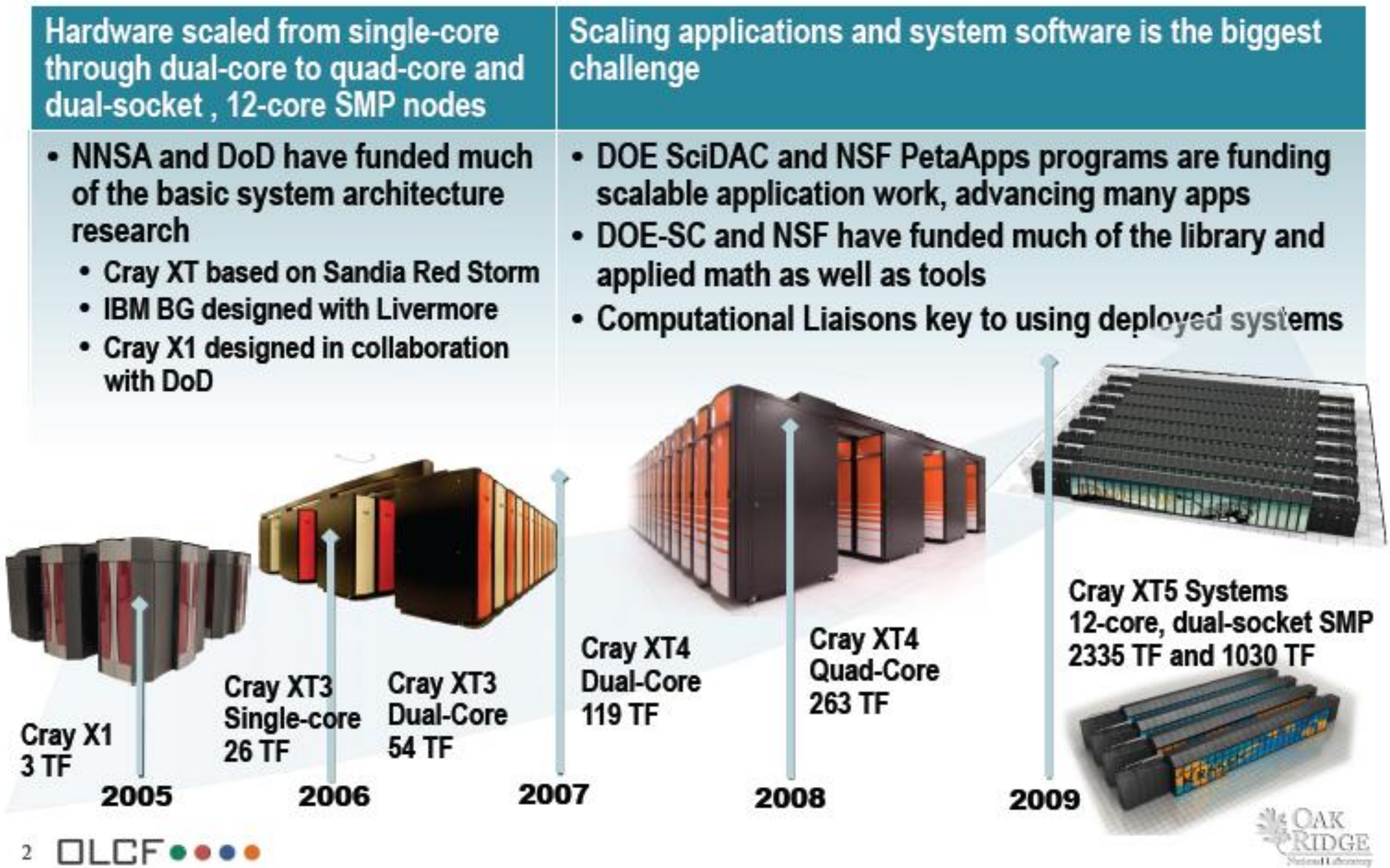
R

Exponential Growth of Computing Power



http://i.timeinc.net/time/daily/2011/1102/singularity_graphic.jpg R

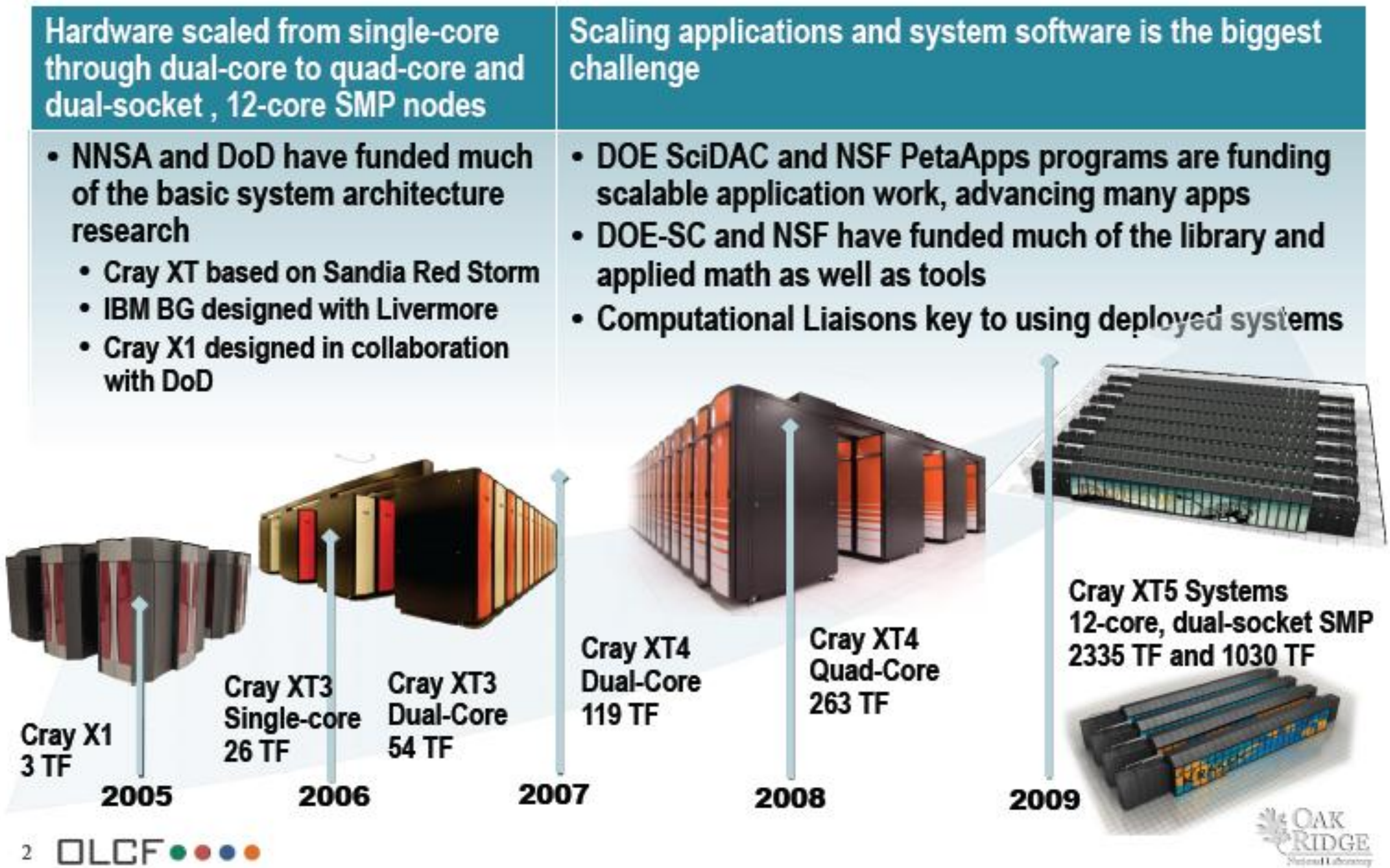
System performance has increased 1000 times since just 2004 at OLCF



Source: B. Messer, ORNL

R

System performance has increased 1000 times since just 2004 at OLCF



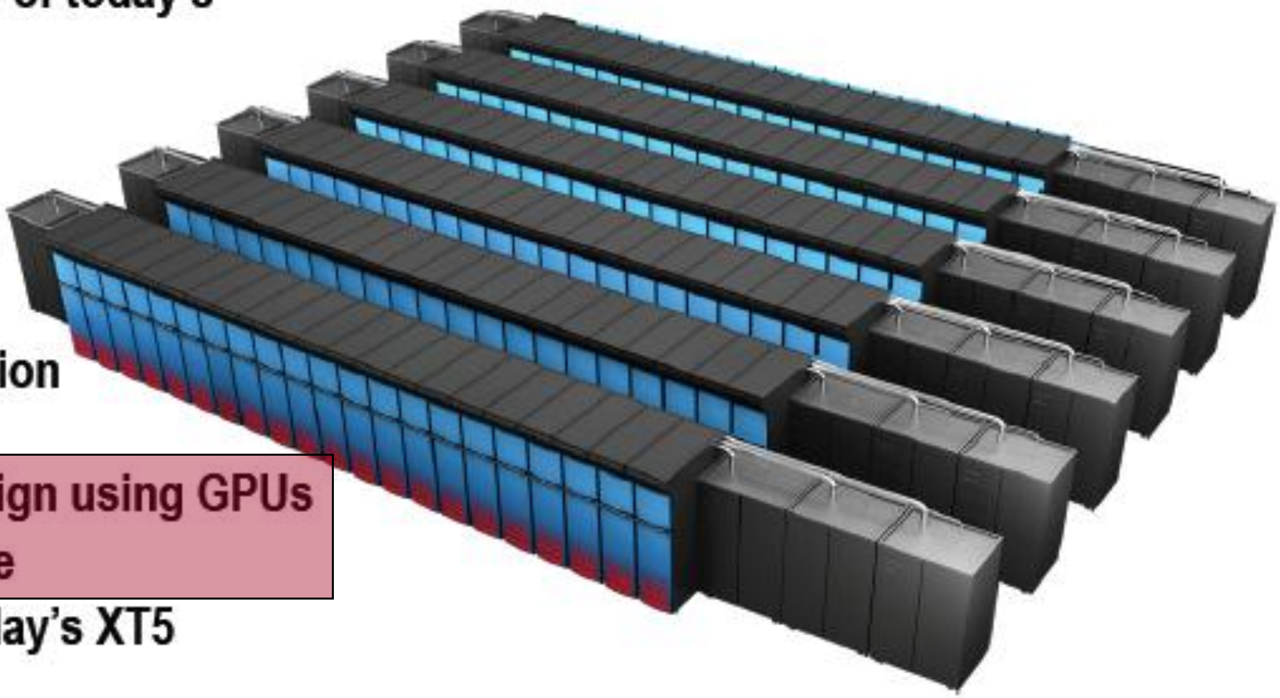
LA-SiGMA machine (condo) at 120TF

Source: B. Messer, ORNL

R

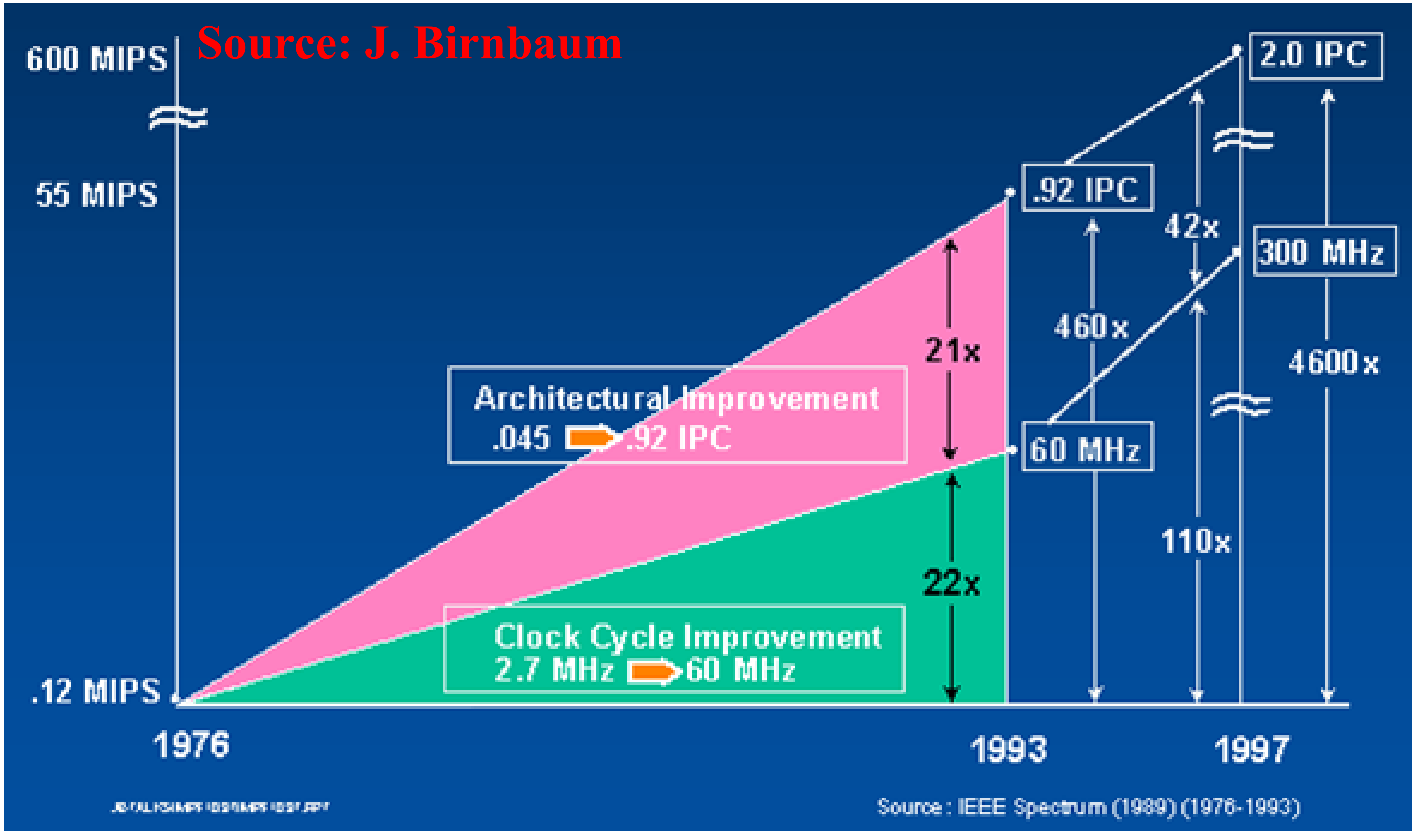
Our near-term target: ORNL's Titan

- Designed for science from the ground up
- Operating system upgrade of today's Linux Operating System
- Gemini interconnect
  - 3-D Torus
  - Globally addressable memory
  - Advanced synchronization features
- New accelerated node design using GPUs
- 10-20 PF peak performance
  - 9x performance of today's XT5
- Larger memory
- 3x larger and 4x faster file system



Source: B. Messer, ORNL

The Good Old Days for Software

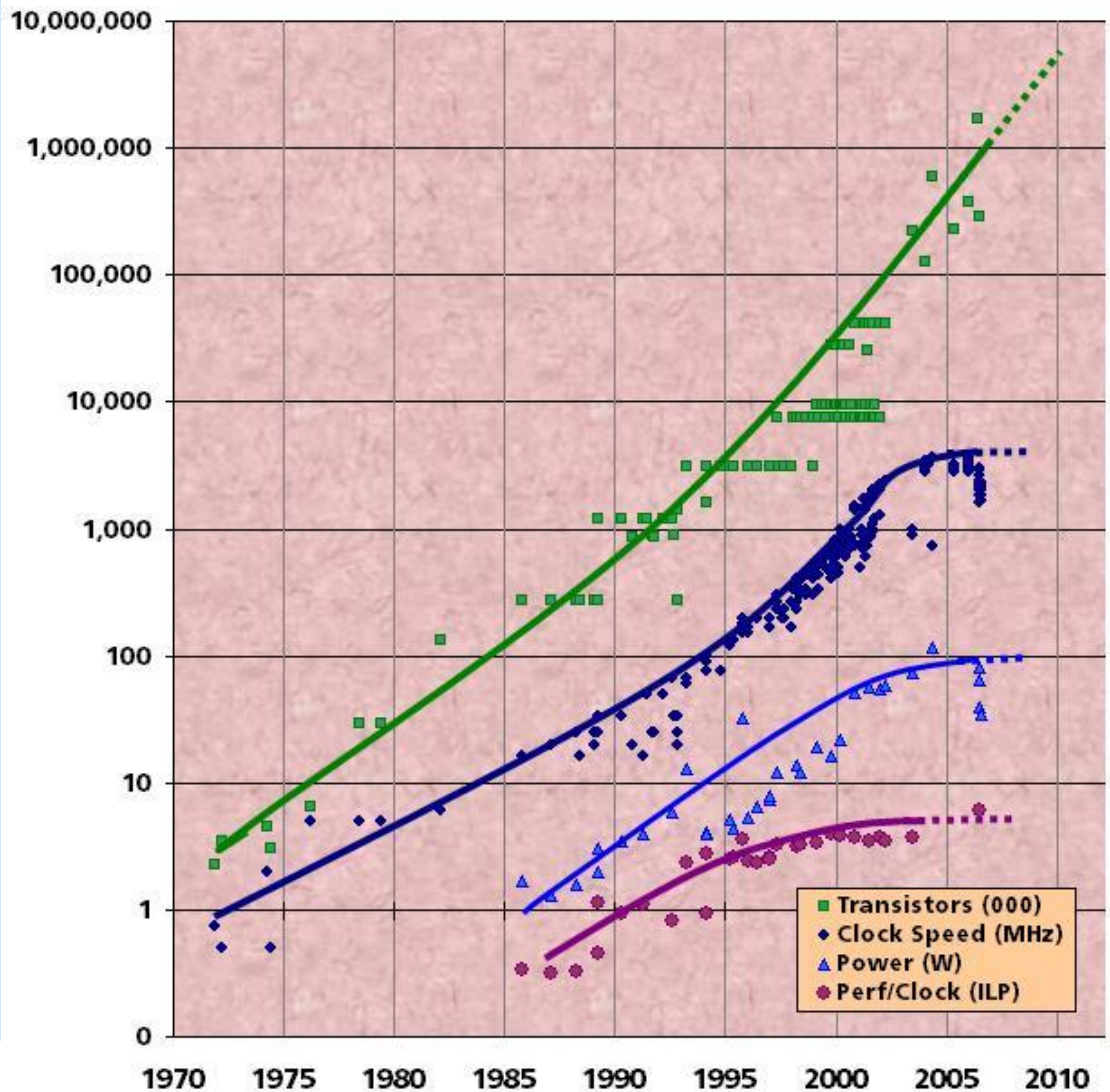


- Single-processor performance experienced dramatic improvements from **clock**, and **architectural** improvement (Pipelining, Instruction-Level-Parallelism)
- Applications experienced automatic performance improvement

R

Revolution is Happening Now

- Chip density is continuing increase ~2x every 2 years
  - Clock speed is not
  - Number of processor cores may double
- There is little or no more hidden parallelism (ILP) to be found
- **Parallelism must be exposed to and managed by software**



Source: Intel, Microsoft (Sutter) and Stanford (Olukotun, Hammond)

R

Challenges to the Future of Supercomputing

- Exaflops-scale computing by 2020
- 100X power consumption improvement
- 1000X increase in parallelism
- 10X improvement in efficiency

R

Challenges to the Future of Supercomputing

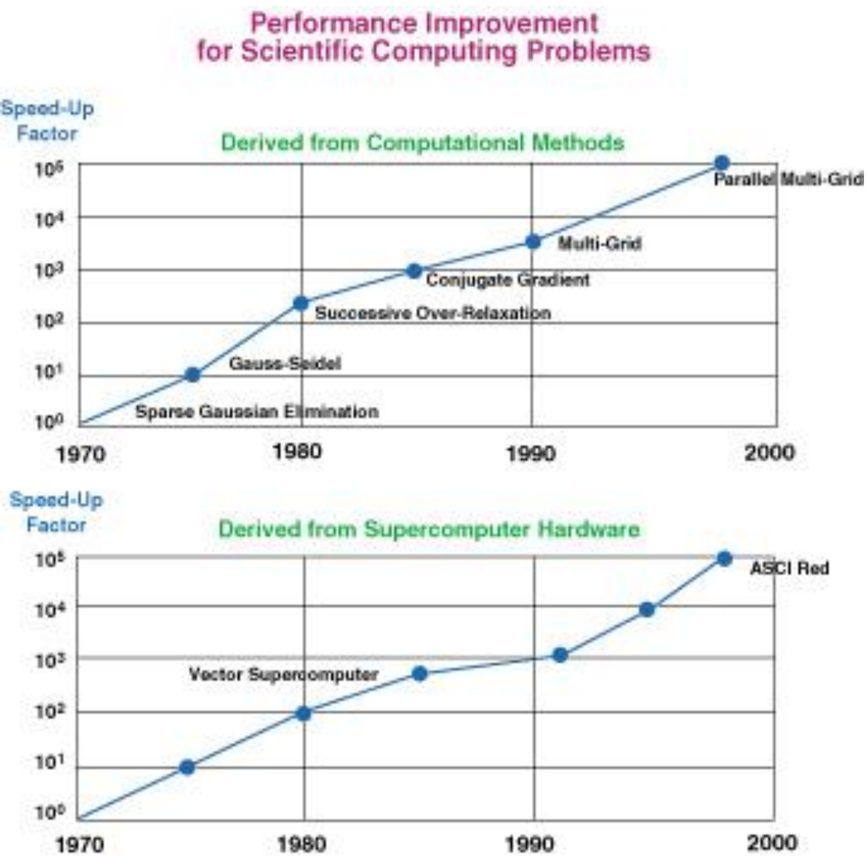
- Exaflops-scale computing by 2020
- 100X power consumption improvement
- 1000X increase in parallelism
- 10X improvement in efficiency
- Exasacle systems will be made up of a mix of multicore and GPU processor architectures
- Improving programmer productivity
- Fault tolerance for reliability

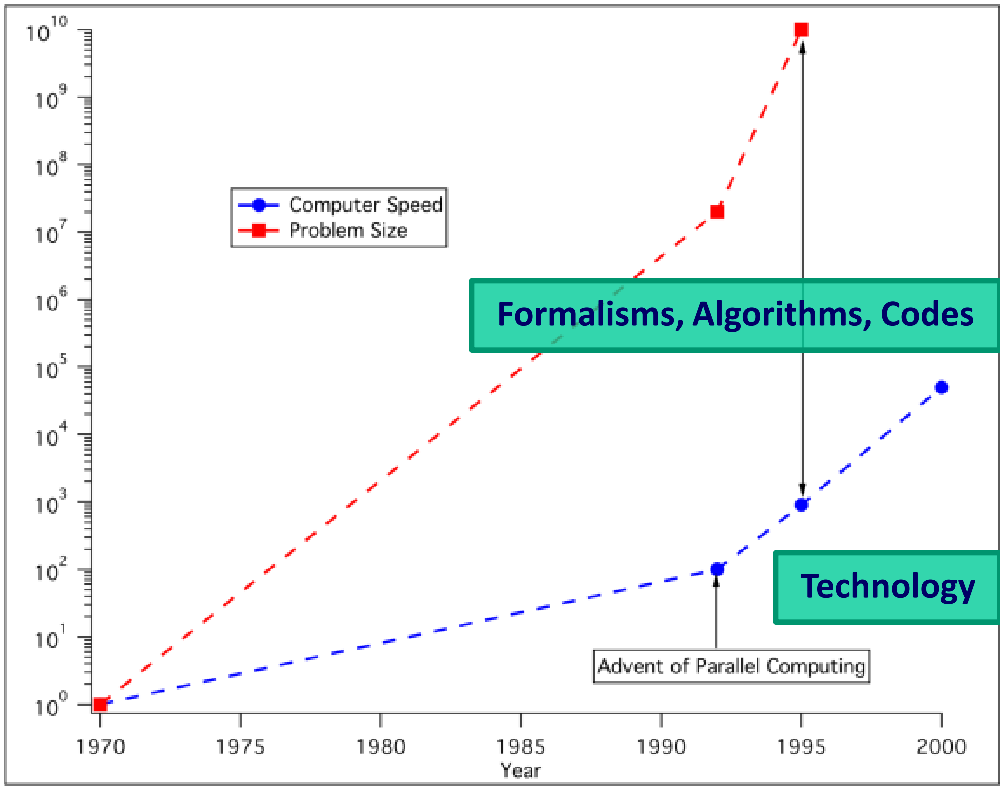
R

Formalisms, Algorithms, Codes, and Technology

Progress in Applied Math

Source: DP Landau



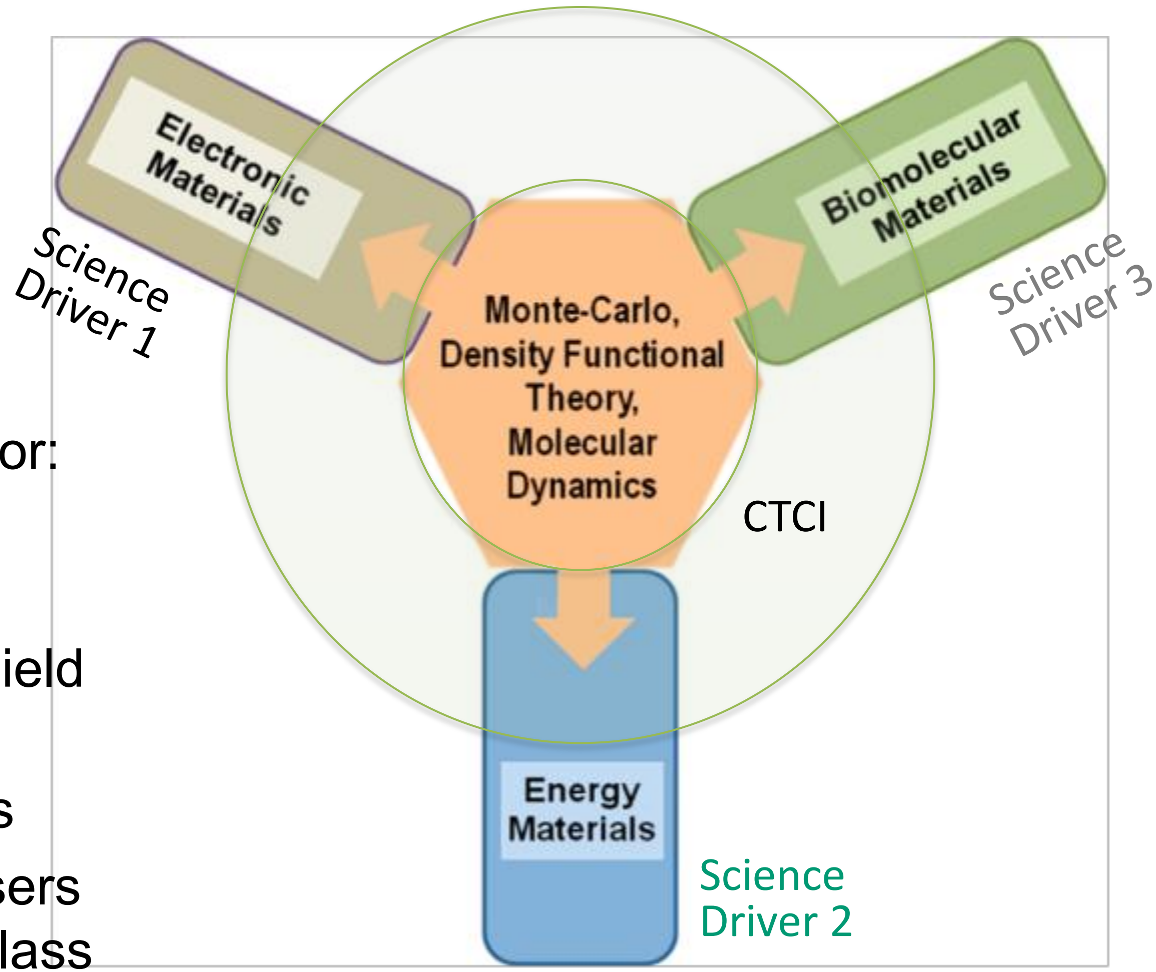


Technology + FAC = Team

M

The Goal

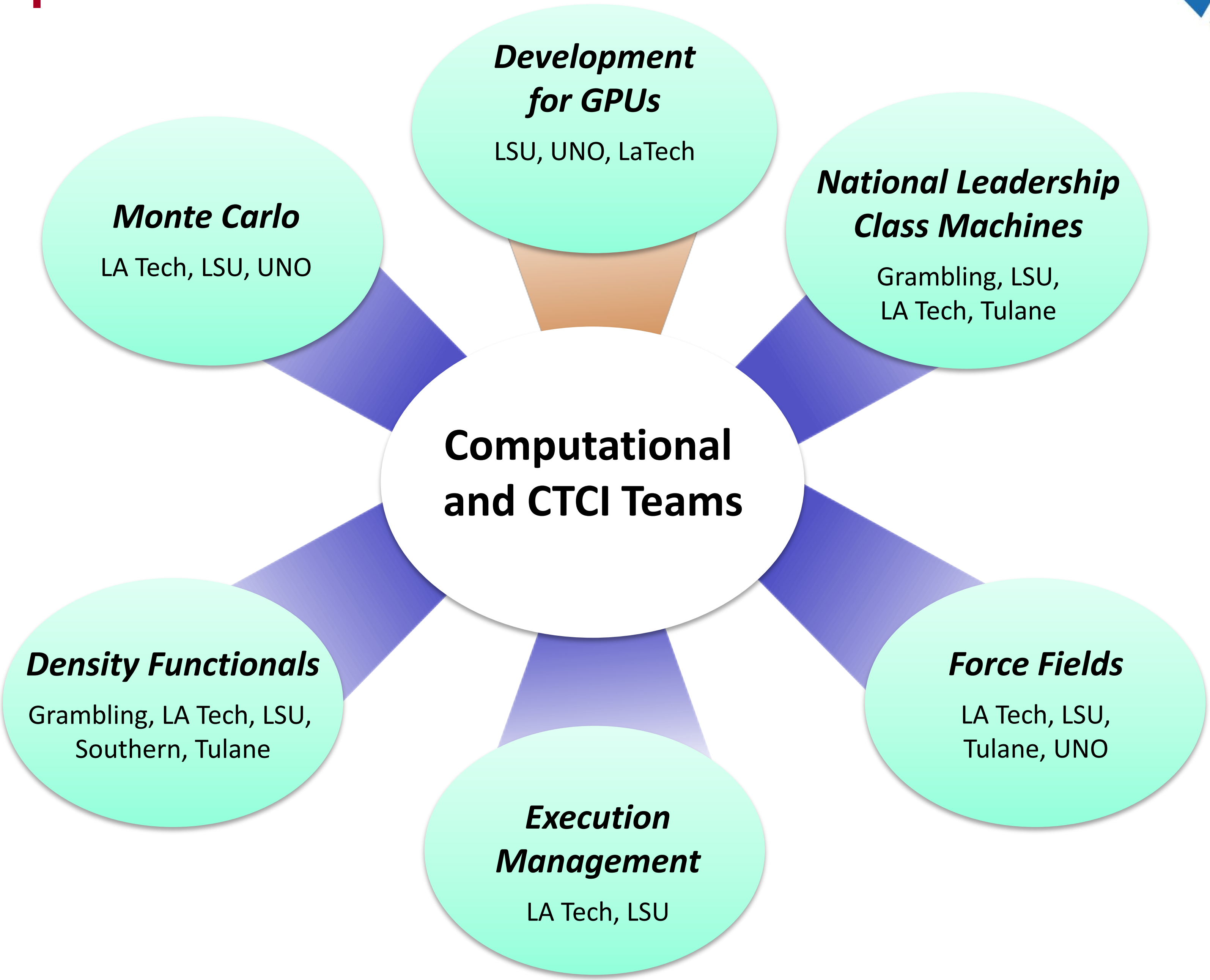
Computational and CTCI Teams:
Build transformational common toolkits



- Build common toolkits for:
  - Monte Carlo
  - Density Functional Theory and Force Field Methods
  - Molecular Dynamics
- Graduate LA-SiGMA users to national leadership class machines

M

Computational/CTCI Research Themes



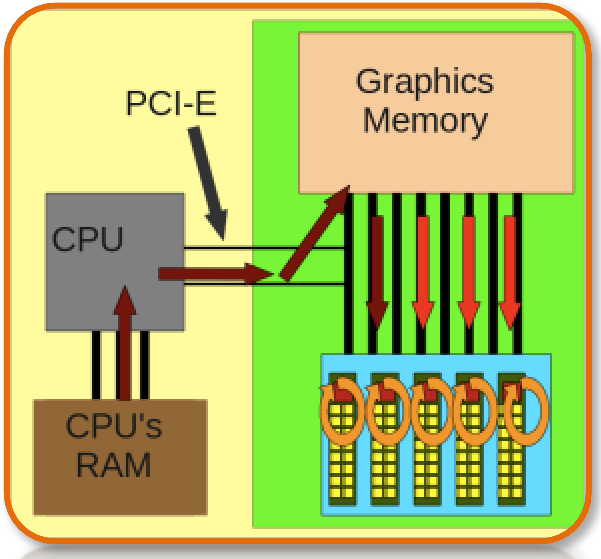
GPU Programming Team

GPU Programming Team

Roughly 25 faculty, students, and postdocs from:

- LSU
- LA Tech
- UNO
- Louisiana School for Math, Sciences, and the Arts (HS)

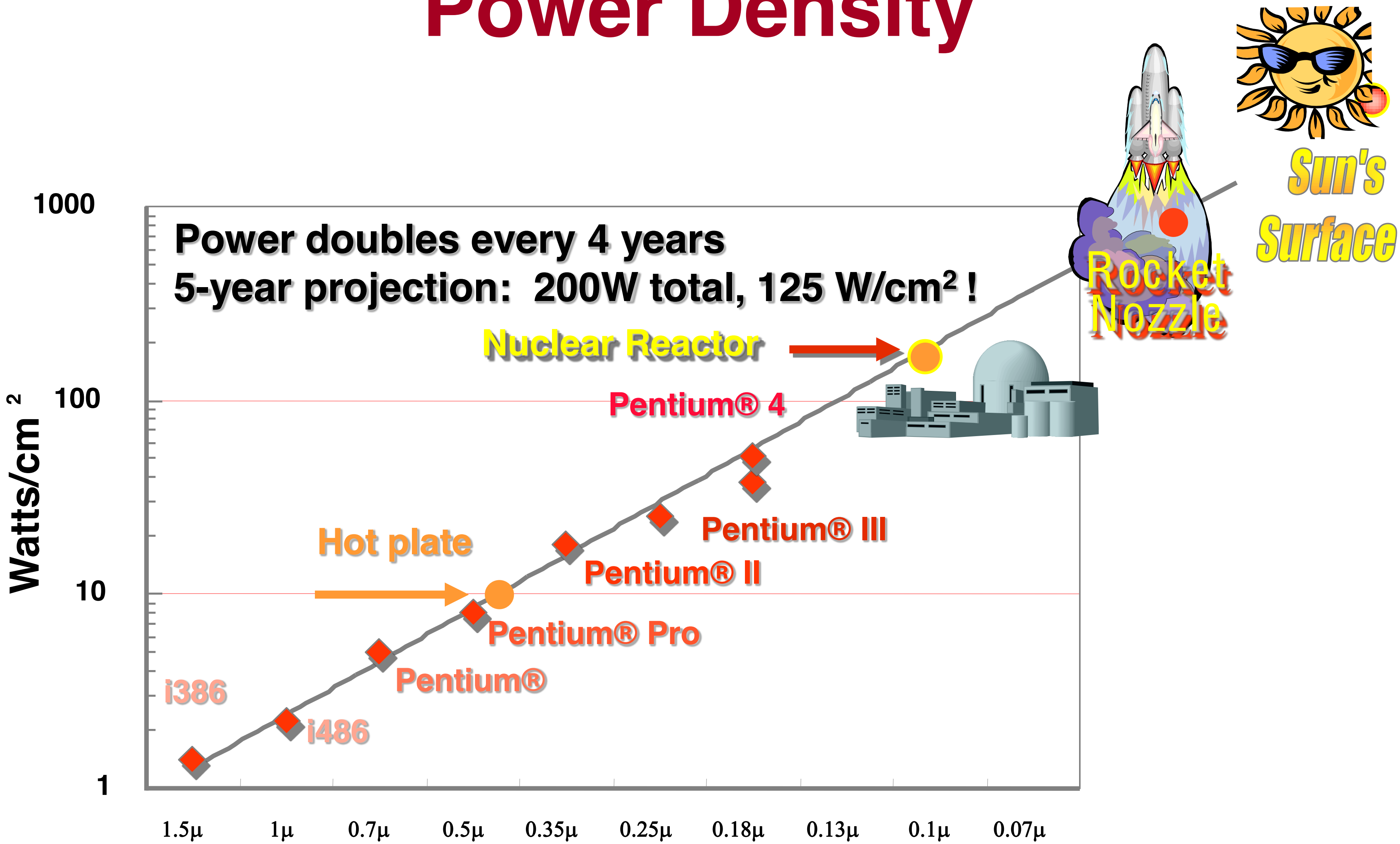
Housed in the "Collaboratorium" at LSU and containing 12 GPU-enabled desktop computers.



LSU "Condo"

LA-SiGMA to purchase GPU-enabled nodes on CCT/LSU's new flagship computer in Fall 2011.

Power Density



* "New Microarchitecture Challenges in the Coming Generations of CMOS Process Technologies" – Fred Pollack, Intel Corp. Micro32 conference key note - 1999. Courtesy Avi Mendelson, Intel.

Moving to Exascale/Green Computing

Jaguar (Cray XT-5): A 2 petaflop machine that draws 10 megawatts of power.

If the increase in computing power is not offset by a reduction in energy use, scaling a Jaguar style machine to an exaflop would take 5 gigawatts of power.

This would require **multiple** nuclear power plants to run

Moving to Exascale/Green Computing

Jaguar (Cray XT-5): A 2 petaflop machine that draws 10 megawatts of power.

If the increase in computing power is not offset by a reduction in energy use, scaling a Jaguar style machine to an exaflop would take 5 gigawatts of power.

This would require **multiple** nuclear power plants to run

Not going to work

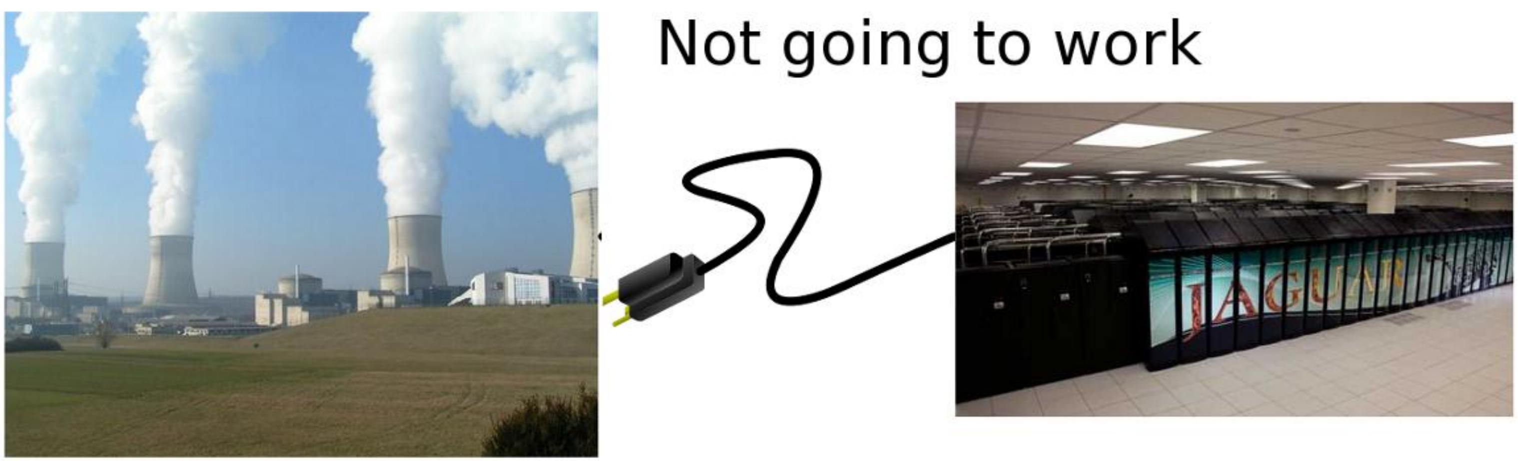
D

Moving to Exascale/Green Computing

Jaguar (Cray XT-5): A 2 petaflop machine that draws 10 megawatts of power.

If the increase in computing power is not offset by a reduction in energy use, scaling a Jaguar style machine to an exaflop would take 5 gigawatts of power.

This would require **multiple** nuclear power plants to run
- A more energy efficient solution isn't just a good idea, it is absolutely necessary to achieve exascale performance



D

Jack Dongarra on GPUs

"GPUs have evolved to the point where many real world applications are easily implemented on them and run significantly faster than on multi-core systems. Future computing architectures will be hybrid systems with parallel-core GPUs working in tandem with multi-core CPUs."

Jack Dongarra

Professor, University of Tennessee

(Top 500, LINPACK, BLAS, …)

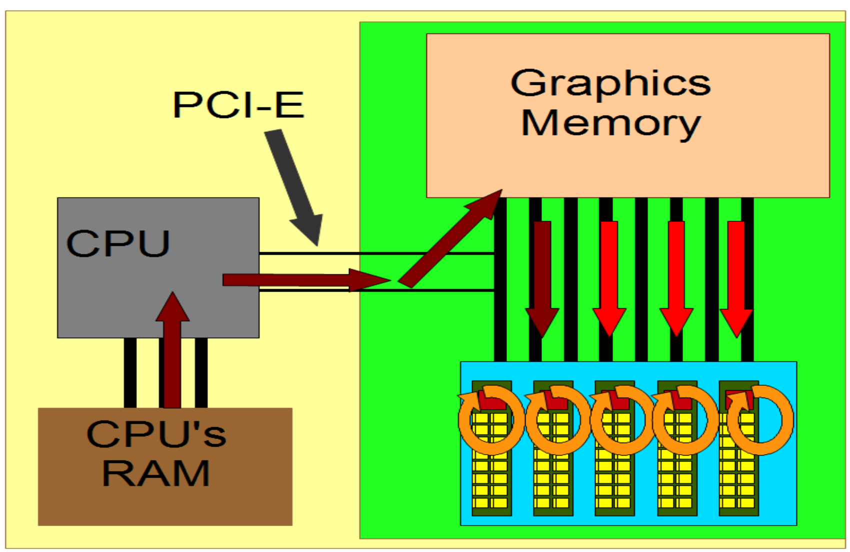
GPU/CPU Differences

CPU

- ~8 Powerful independent processors
- ~16 GB RAM, 256K cache per processor
- Large power requirement
- ~0.17 TFLOPs

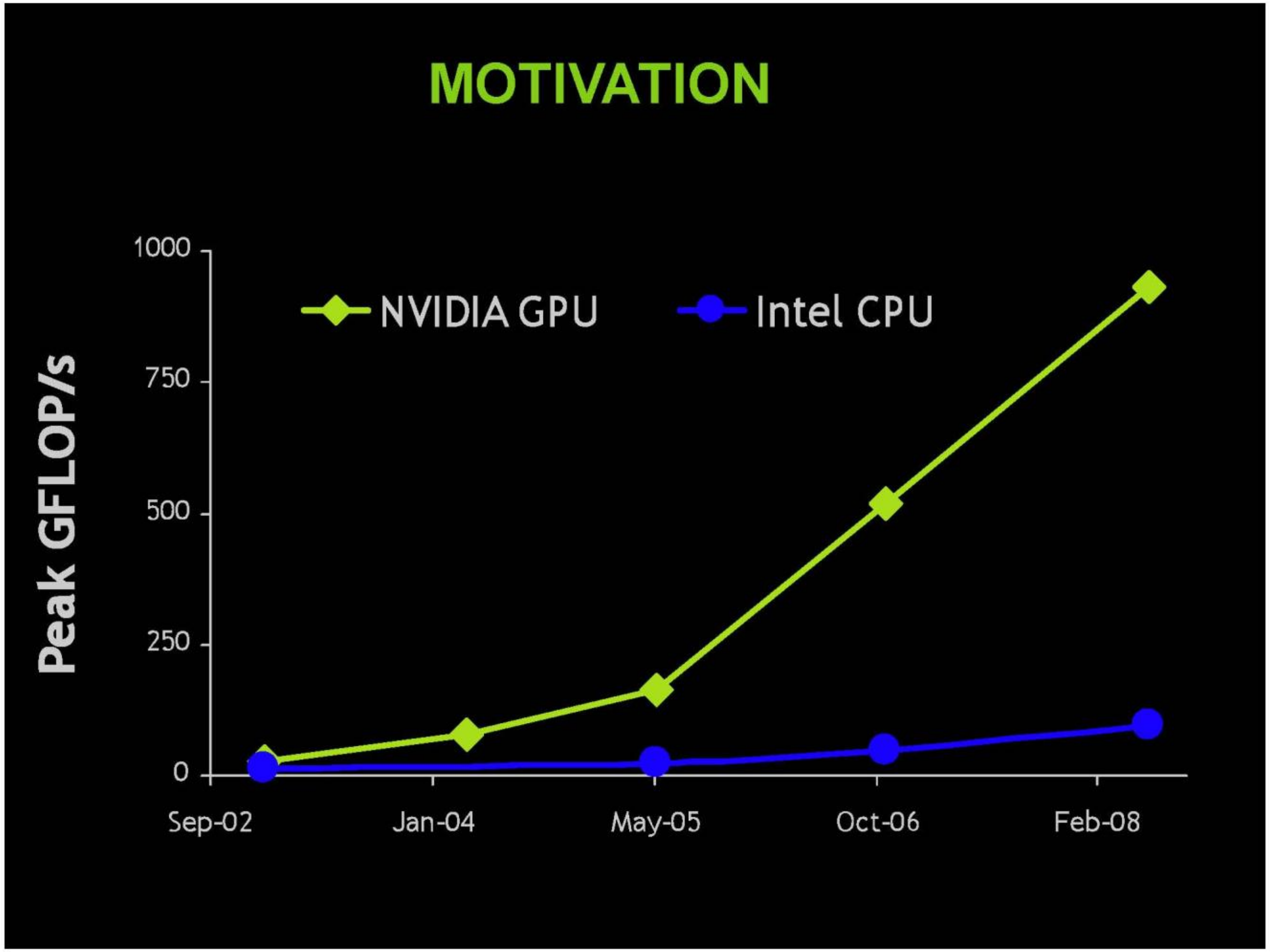
GPU

- ~512 cores in sets of 32 or 48 operating in lockstep
- ~12 MB RAM, 2KB cache per processor
- Low power requirement
- >1.5 TFLOPs
- No virtual memory



D

Graphics Processing Units (GPUs)



Source: NVIDIA

GPU applications to SD mapping

Projects	SD mapping
Hirsch-Fye Monte Carlo	SD1
Parallel Tempering	SD1, SD2, SD3
Exact Diagonalization (Lanczos)	SD1
Variational Monte Carlo	SD1
Coupled FDTD-Micromagnetic-Landau-Devonshire	SD1
Grand Canonical Monte Carlo	SD2, SD3

GPU applications: Challenges

- More processors but less memory per processor compared to CPUs
- Requires extraction and management of parallelism: lots of threads
- Threads are parallel and perform the same computation on different data
- Manage slow data movement between CPU and GPU (separate GPU and CPU memories)
- Hard resource limits: your computation will not run on a GPU if you exceed resource limits

Needs a team approach

D

GPU: Parallel Programming Tools

- Collaboration: Leungsuksun, Drs. Alex Safarenko and Sven-Bodo Scholz, University of Hertfordshire, UK.
- Parallel Programming Tool Development based on SAC (Single Assignment C) toolset that enables parallel application developers expressing their problems in a high-level language
- Tool set is capable to generate codes in various HPC architectures including multicore and GPU.
- The collaborative enhancement will introduce fault tolerance to SAC and its parallel applications.

GPU: Compilation Technology

- GPUs are cost-effective and powerful processors for general-purpose computing
- Programming GPUs (for example, the NVIDIA GPUs using CUDA or OpenCL) is still tedious:
  - The performance of the GPU is highly sensitive to the formulation of the kernel; needs significant experimentation
  - Programmers may like this low level of control and this is appropriate for library development; unlike the case of general-purpose processors, compilers and tools are not quite helpful here
- **Strategy:** Develop a system (**PLUTO**) for automatic transformation of sequential input programs into efficient parallel CUDA/OpenCL programs

R

Graduating to National Leadership Class (NLC) Machines: a CTCI Goal

- Explore parallelism and scalability
- Get experience with code development on smaller clusters such as LONI/teragrid
- Demonstrate how your codes will scale to the NLC machines (partnering with national labs)
  - How do you do this when you do not yet have access to the machine?
- Apply for compute time on NLC machines
  - Example: INCITE proposal (with PNNL)
  - Example: Continuing work on Kraken

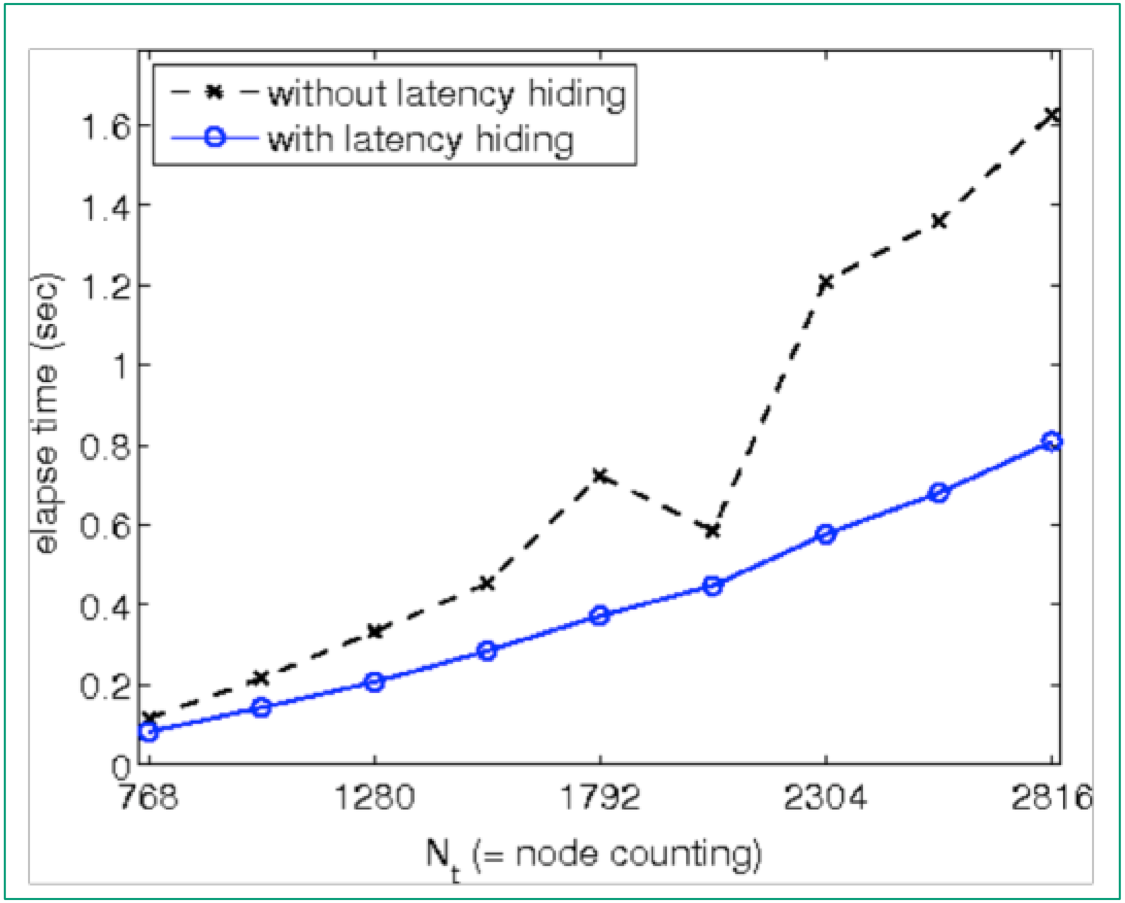
Scaling many-body codes to 30,000 processors

- Multi-Scale Many Body (MSMB) methods are required to more accurately model competing phases, including magnetism, superconductivity, insulating behavior etc. of strongly-correlated materials
  - In these multi-scale approaches, explicit methods are used only at the shortest length scales.
  - Two particle many-body methods, which scale algebraically, are used to approximate the weaker correlations at intermediate length scales.

R

Scaling of Multi-Scale Methods to 30000 Cores

Latency hiding techniques, that overlap the communication with local computations, allow efficient use of 30,000 or more processors at NSF supercomputing facilities.



K.M. Tan et al. (in preparation, 2011)

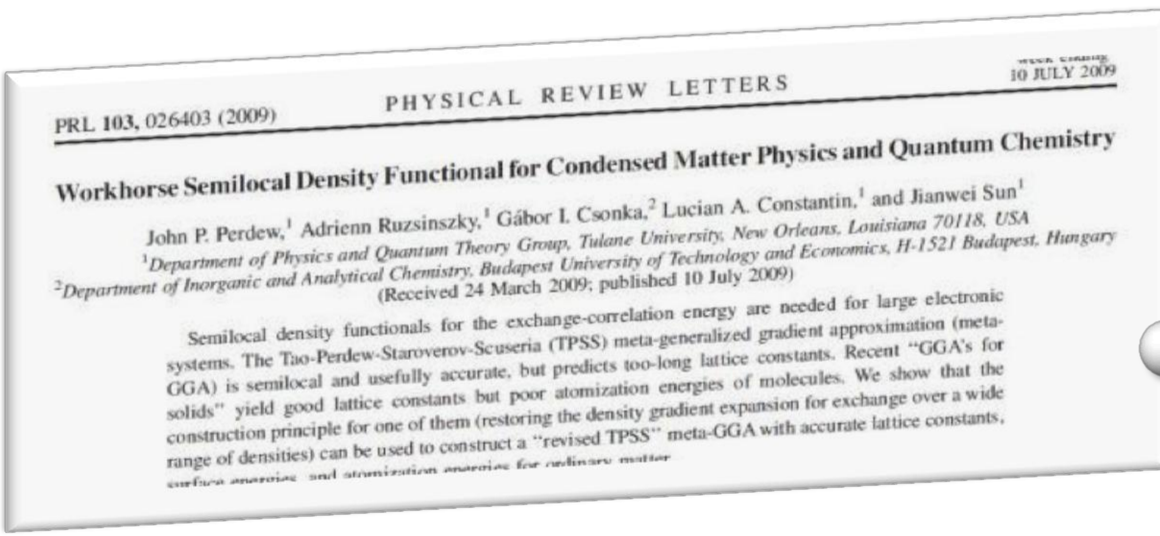
R

Density Functionals and Force Fields

Perdew group has developed a “work-horse semilocal functional” [*Phys. Rev. Lett*. **103**, 026403 (2009)] for large electronic systems that yields accurate lattice constants, surface energies, and atomization energies.

This functional has been incorporated into VASP, a massively parallel DFT code

We are constructing force fields combining *ab initio* calculations of small clusters with different DFT functionals and bulk simulations/calculations.


PRL 103, 026403 (2009) PHYSICAL REVIEW LETTERS 10 JULY 2009

Workhorse Semilocal Density Functional for Condensed Matter Physics and Quantum Chemistry

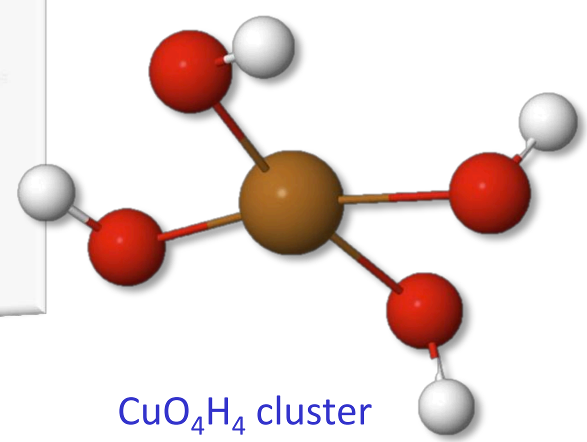
John P. Perdew,[1] Adrienn Ruzsinszky,[1] Gábor I. Csonka,[2] Lucian A. Constantin,[1] and Jianwei Sun[1]

[1]*Department of Physics and Quantum Theory Group, Tulane University, New Orleans, Louisiana 70118, USA*

[2]*Department of Inorganic and Analytical Chemistry, Budapest University of Technology and Economics, H-1521 Budapest, Hungary*

(Received 24 March 2009; published 10 July 2009)

Semilocal density functionals for the exchange-correlation energy are needed for large electronic systems. The Tao-Perdew-Staroverov-Scuseria (TPSS) meta-generalized gradient approximation (meta-GGA) is semilocal and usefully accurate, but predicts too-long lattice constants. Recent “GGA's for solids” yield good lattice constants but poor atomization energies of molecules. We show that the construction principle for one of them (restoring the density gradient expansion for exchange over a wide range of densities) can be used to construct a “revised TPSS” meta-GGA with accurate lattice constants,

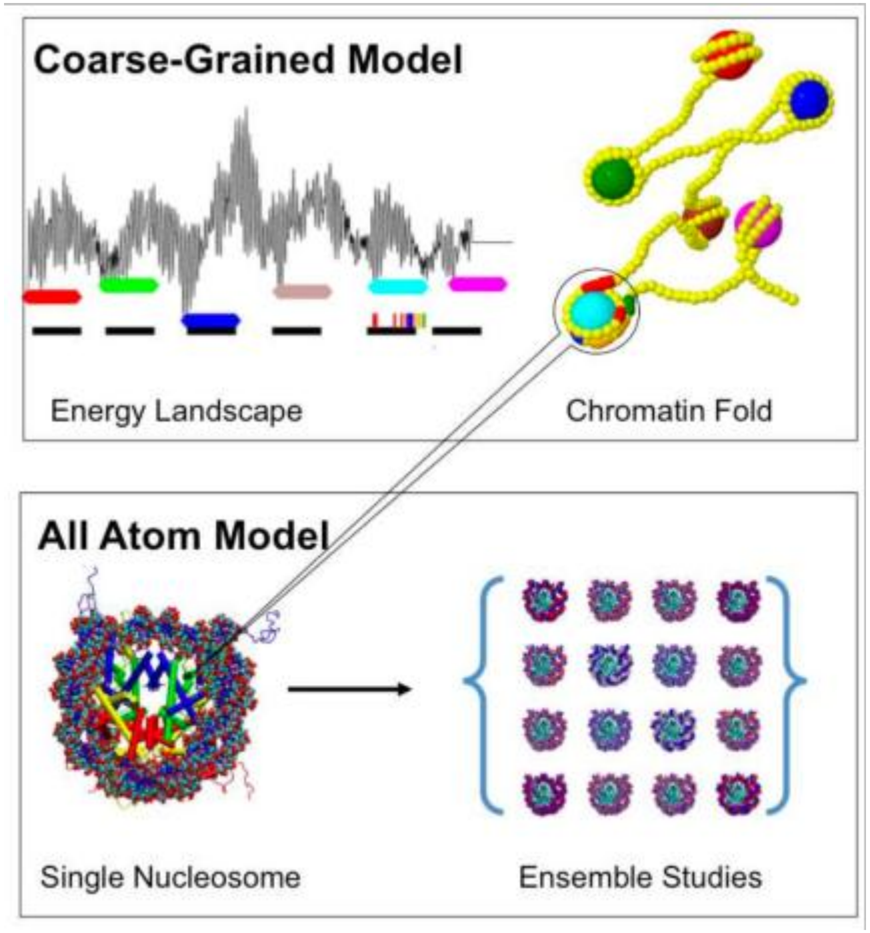


CuO_4H_4 cluster

LSU (Hall, Dellinger),
La Tech (Wick, Ramachandran)

DFT/Force Field Teams

Molecular Dynamics and Execution Management

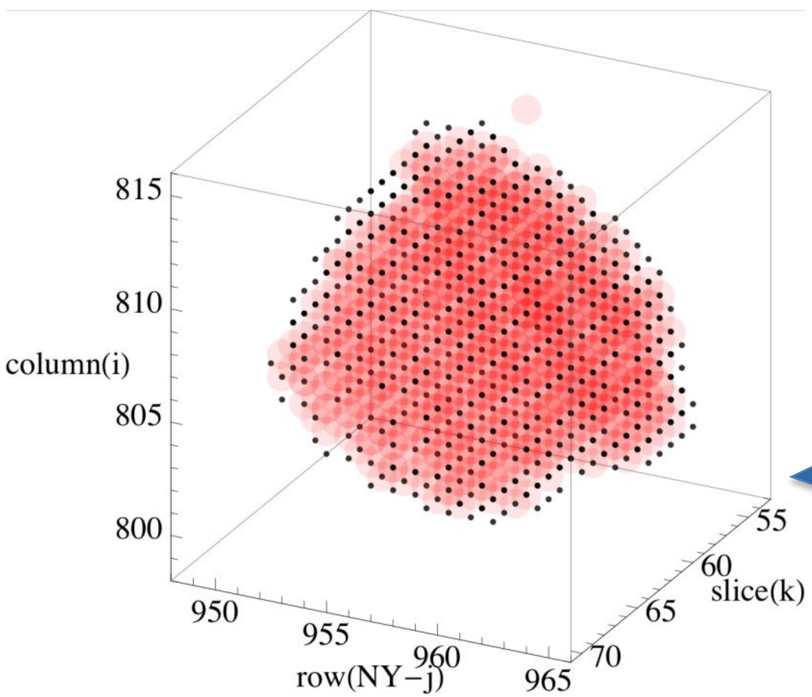


LA Tech (Bishop), LSU (Jha)

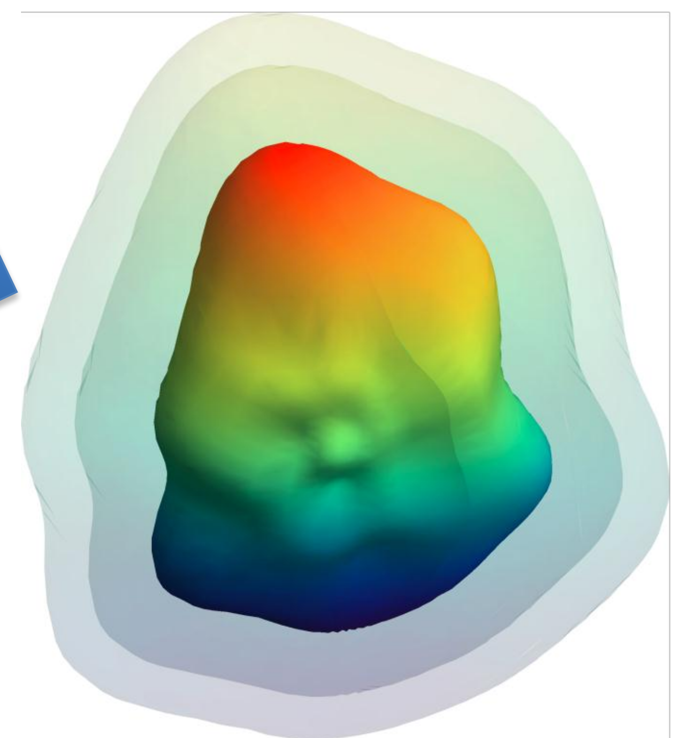
Efficiently distributed molecular simulations of solvated nucleosomes combining coarse-grained models of the folding of DNA with all atom models.

~6,000 simulation tasks associated with this project.

Experimental Inverse Problem



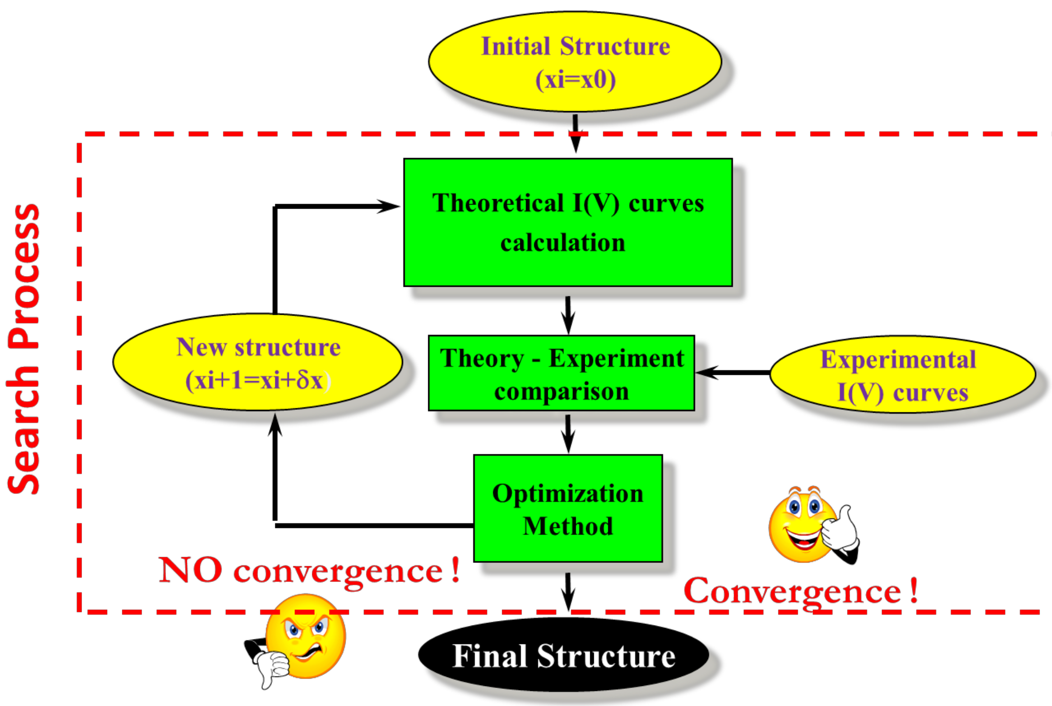
Example: From X-ray and neutron scattering data to images of flame retardants in polymer blends



Using the results of actual observations to infer the values of the parameters characterizing the system under investigation

Inverse Problem in Structure Determination by LEED

- Multiple scattering forces LEED analysis to be indirect
- Quantitative comparison theory-experiment
- Search for the best fitted structural model: Search Problem
- Hard task: locate the global minimum in an N-dimensional parameter space;



Complex Transition Metal Oxides :
complex structure –> many structural parameters to be optimized;

Necessary Improvements
1) Global Search Methods;
2) Faster Multiple Scattering calculations
3) **Direct Methods (inverse problem),** surface structure directly from experimental data -> **Holy Grail of LEED !**

Appreciate help and suggestions for team effort

M

ParalleX: Runtime Support for Extreme Scale Execution

- "*MPI must evolve or die*" (Al Geist, ORNL)
  Current models such as MPI and Open MP are not expected to work well on exascale systems
- ParalleX: A new execution model and runtime system that attacks problems of starvation, latency, and overhead
- An experimental proof of concept software (HPX) is available for ParalleX; application experiments demonstrate 2-3X performance gain
- Plan to use HPX with LA-SiGMA codes; this experience will shape the evolution of ParalleX

R

Temporal Pattern Discovery in MD simulation data

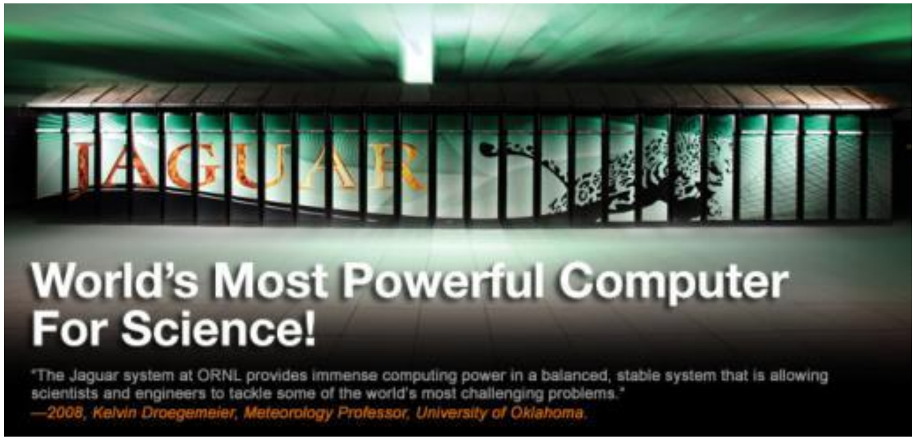
- Multivariate temporal data are collections of contiguous data values that reflect complex temporal changes over a given duration.
- Traditional data mining techniques do not utilize the complex behavioral changes in MD simulation data over a period of time and are not well suited to classify multivariate temporal data
- Developed an automated system that builds a temporal pattern-based framework to classify multivariate time series data; results show that this method produces overall superior results.

Funding and Outreach

- INCITE proposal (in collaboration with Pacific Northwest National Laboratories) for compute cycles on Jaguar and Titan
- NSF SISI proposal for GPU code development
- Indo-US Center (IUSSTF)
- SCiDAC proposal under development
- Outreach:
  - Summer REU and RET programs (2011)
  - Beowulf Boot Camp for High School Students and Teachers, 2011
  - Conference tutorials on GPUs: (i) International Symp. on Code Gen. & Opt., April 2011; (ii) Intl. Conf. Supercomputing, June 2011

Partnership with Pacific Northwest National Lab

- Environmental Molecular Sciences Laboratory (EMSL)
  - Home of NWChem!
- DOE INCITE request for 200,000,000 SUs over 3 years on Jaguar and Titan supercomputers (20+ PFLOPs, Kepler GPUs)
- Internship program for LA-SiGMA students
  - 3-6 month visits
  - Working with EMSL Open Source Code Developers
  - Opportunities for students, undergraduate to PhD
- Partners in Indo-US SCES Centre

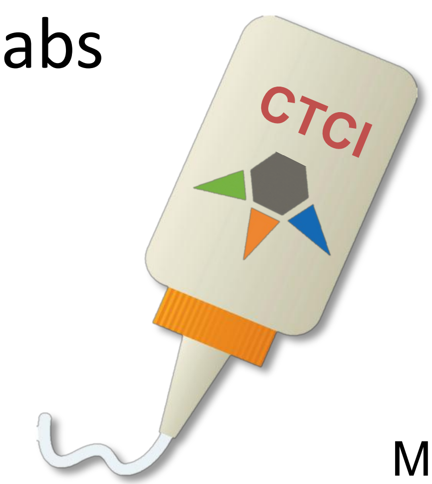


EMSL Associate Lab Director Bill Shelton participating in an LA-SiGMA REU panel

Summary

- CTCI: Leveraging the exponential increase in computer power
  - Graduating our users to current national leadership class machines
  - Preparing users for next-generation computers
  - Developing common computational toolkits
  - Expanding collaborations within LA-SiGMA and developing partnerships with national labs

"The glue" that binds the different SDs



M

Thank You