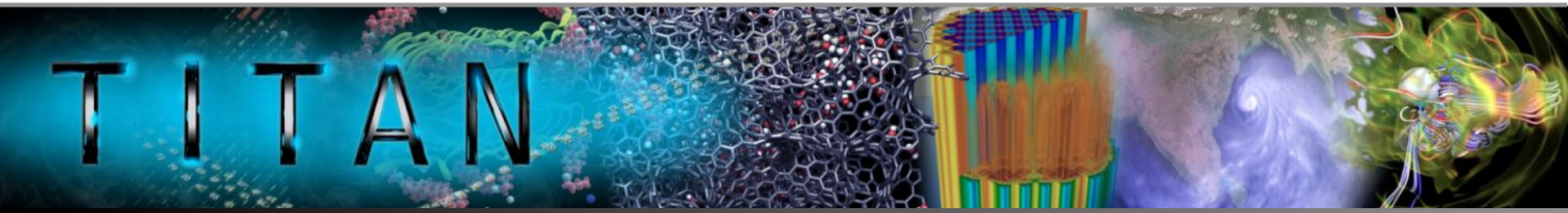
What does Titan tell us about preparing for exascale supercomputers?

20 Years of Excellence in Computational Science
OLCF
Oak Ridge Leadership Computing Facility
1992–2012

Jack Wells
Director of Science
Oak Ridge Leadership Computing Facility

Programme du 32eme Forum ORAP
Maison de la Simulation, Saclay, October 10, 2013



Abstract

Modeling and simulation with petascale computing has supercharged the process of innovation and understanding, dramatically accelerating time-to-insight and time-to-discovery. This presentation will focus on early outcomes from the Titan supercomputer at the Oak Ridge National Laboratory. Titan has over 18,000 hybrid compute nodes consisting of both CPUs and GPUs. In this presentation, I will discuss the lessons we have learned in deploying Titan and preparing applications to move from conventional CPU architectures to a hybrid machine. I will present early results of applications running on Titan and the implications for the research community as we prepare for exascale supercomputer in the next decade. Lastly, I will provide an overview of user programs at the Oak Ridge Leadership Computing Facility with specific information how researchers may apply for allocations of computing resources.

Outline

- U.S. DOE Leadership Computing Program
  - –INCITE & other user programs
- Hardware Trends: Increasing Parallelism
- OLCF-3: The Titan Project
- Application readiness and early results on Titan
- OLCF-4: Getting ready for our next machine in 2017

What is the Leadership Computing Facility (LCF)?

- Collaborative DOE Office of Science program at ORNL and ANL
- Mission: Provide the computational and data resources required to solve the most challenging problems.
- 2-centers/2-architectures to address diverse and growing computational needs of the scientific community
- Highly competitive user allocation programs (INCITE, ALCC).
- Projects receive 10x to 100x more resource than at other generally available centers.
- LCF centers partner with users to enable science & engineering breakthroughs (Liaisons, Catalysts).

OAK RIDGE National Laboratory
Argonne NATIONAL LABORATORY

Big Problems Require Big Solutions

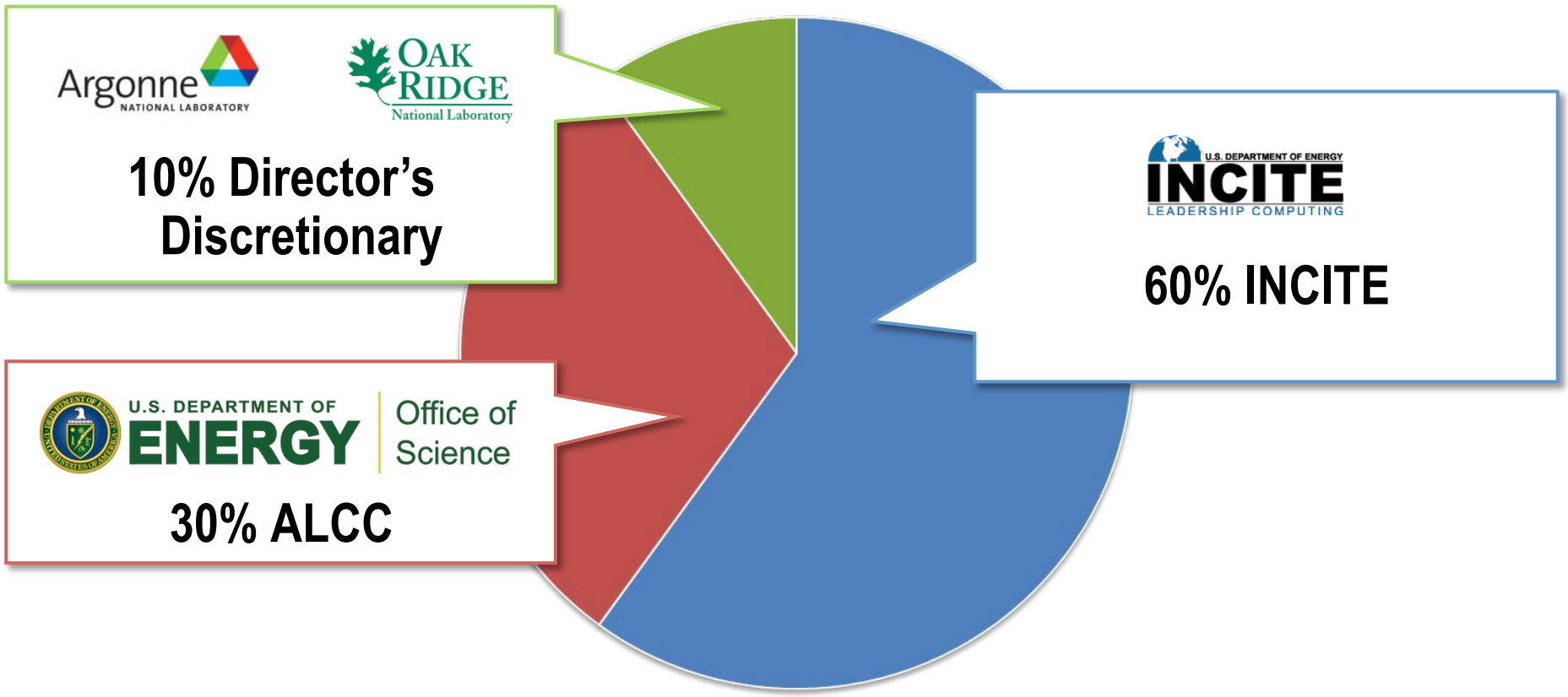
Climate Change

Energy

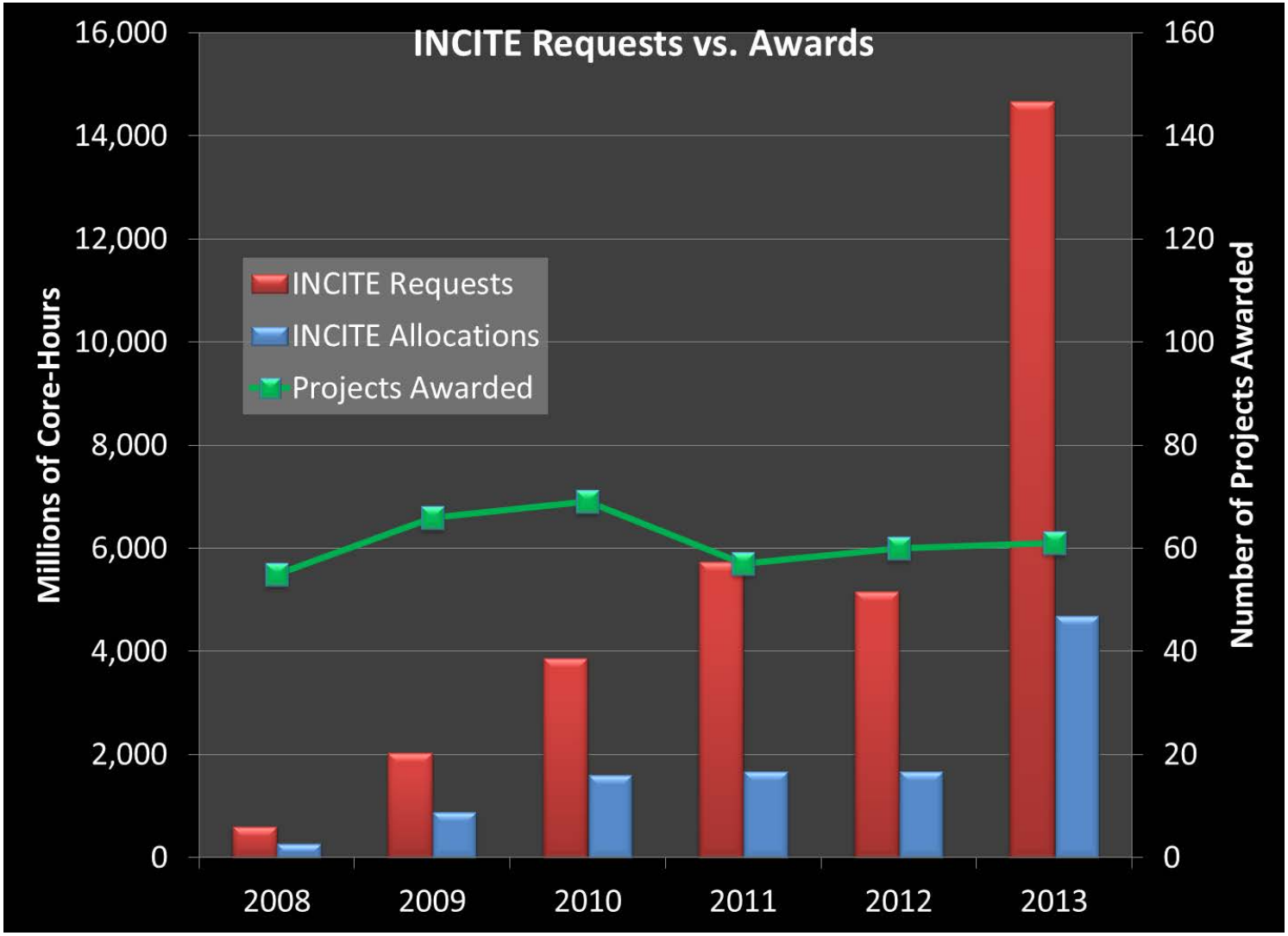
Healthcare

Competitiveness

LCF User Programs:
More than 6.5 billion core hours awarded in 2013

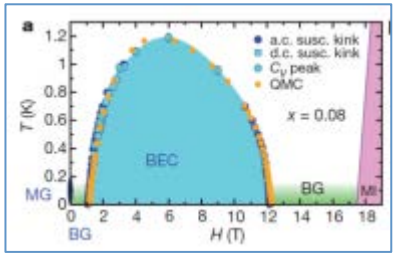


Demand for INCITE resources outstrips supply with 3x more time requested than available – Number of projects remains flat



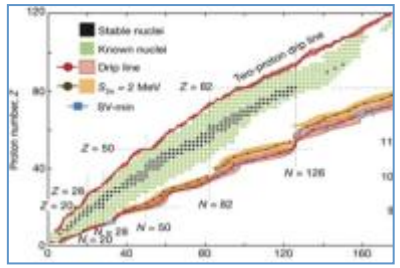
High-impact science done at OLCF across a broad range of disciplines.

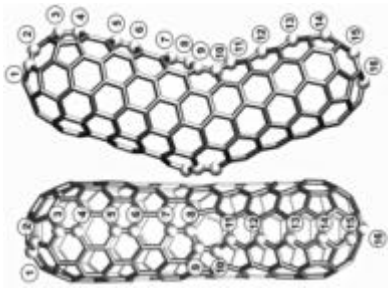
For example in 2012:



Materials: Quantum Magnets
"Bose glass and Mott glass of quasiparticles in a doped quantum magnet"
Rong Yu (Rice U.)
Nature (2012)

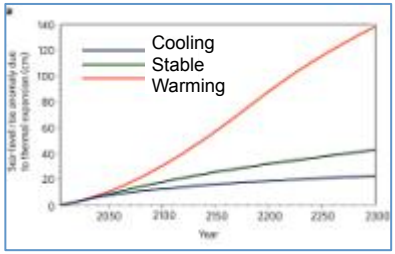
Nuclear Physics
"The Limits of the Nuclear Landscape"
J. Erier, (UT/ORNL)
Nature (2012)

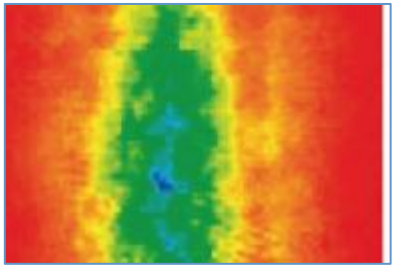




Carbon Nanomaterials
"Covalently bonded three-dimensional carbon nanotube solids via boron induced nanojunctions"
Hashim (Rice), **Scientific Reports** (2012)

Climate Prediction and Mitigation
"Relative outcomes of climate change mitigation related to global temperature versus sea-level rise"
G.A. Meehl (NCAR),
Nature Climate Change (2012)

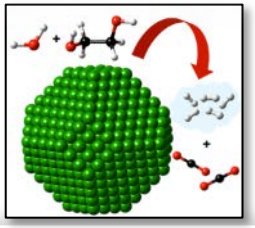
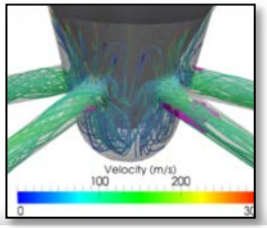
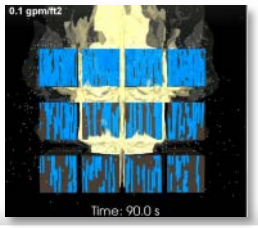
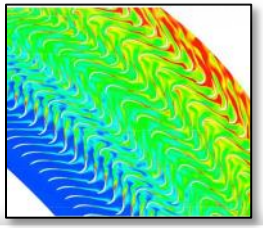
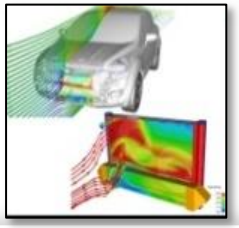




Plasma Physics:
"Dynamics of relativistic transparency and optical shuttering in expanding overdense plasmas"
S. Palaniyappan (LANL)
Nature Physics (2012)

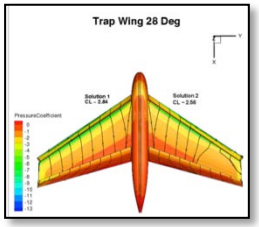
Paleoclimate Climate Change
"Global warming preceded by increasing carbon dioxide concentrations during the last deglaciation"
J. Shakun, (Harvard/Columbia)
Nature (2012)

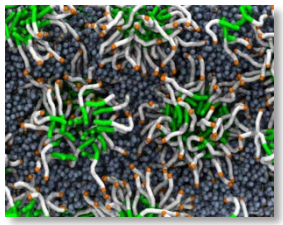
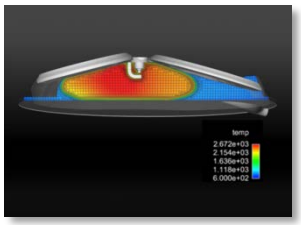
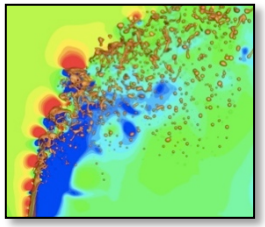
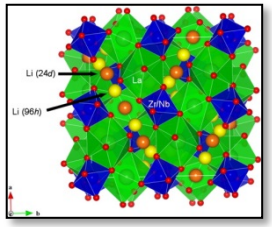
Innovation through Industrial Partnerships

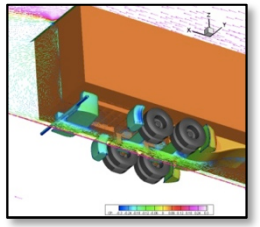
United Technologies Research Center	RAMGEN POWER SYSTEMS	GM	FM Global	GE	Ford
Catalysis	**Design innovation**	**Gasoline engine injector**	**Industrial fire suppression**	**Turbo machinery efficiency**	**Underhood cooling**
Demonstrated biomass as a viable, sustainable feedstock for hydrogen production for fuel cells; showed that nickel is a feasible catalytic alternative to platinum	Accelerating design of shock wave turbo compressors for carbon capture and sequestration	Optimization of injector hole pattern design for desired in-cylinder fuel-air mixture distributions (4−40× potential improvement in workflow throughput via 100s of ensemble simulations)	Developing high-fidelity modeling capability for fire growth and suppression; fire losses account for 30% of U.S. property loss costs	Simulated unsteady flow in turbo machinery, opening new opportunities for design innovation and efficiency improvements.	Developed a new, efficient and automatic analytical cooling package optimization process leading to one of a kind design optimization of cooling systems
					

Innovation through Industrial Partnerships

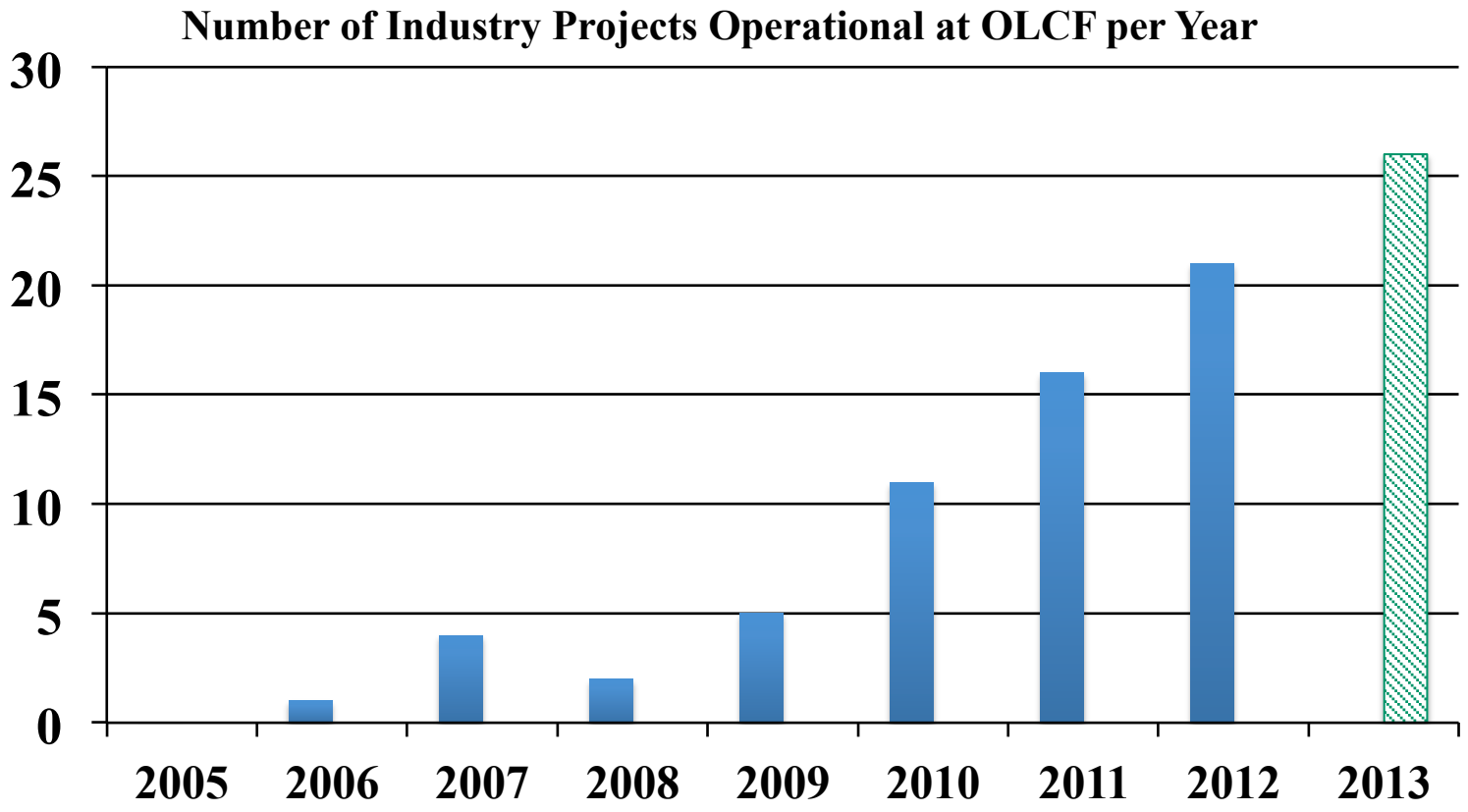
Aircraft design	Consumer products	Engine cycle-to-cycle variation	Jet engine efficiency	Li-ion batteries	Long-haul truck fuel efficiency
Simulating takeoff and landing scenarios improved a critical code for estimating characteristics of commercial aircraft, including lift, drag, and controllability	Leadership computing and molecular dynamics software advanced understanding of chemical processes that can limit product shelf life	Emerging model of engine cyclic variation will apply thousands of processors to a challenging problem	Accurate predictions of atomization of liquid fuel by aerodynamic forces enhance combustion stability, improve efficiency, and reduce emissions	New classes of solid inorganic Li-ion electrolytes could deliver high ionic and low electronic conductivity and good electrochemical stability	Simulations reduced by 50% the time to develop a unique system of add-on parts that increases fuel efficiency by 7–12%





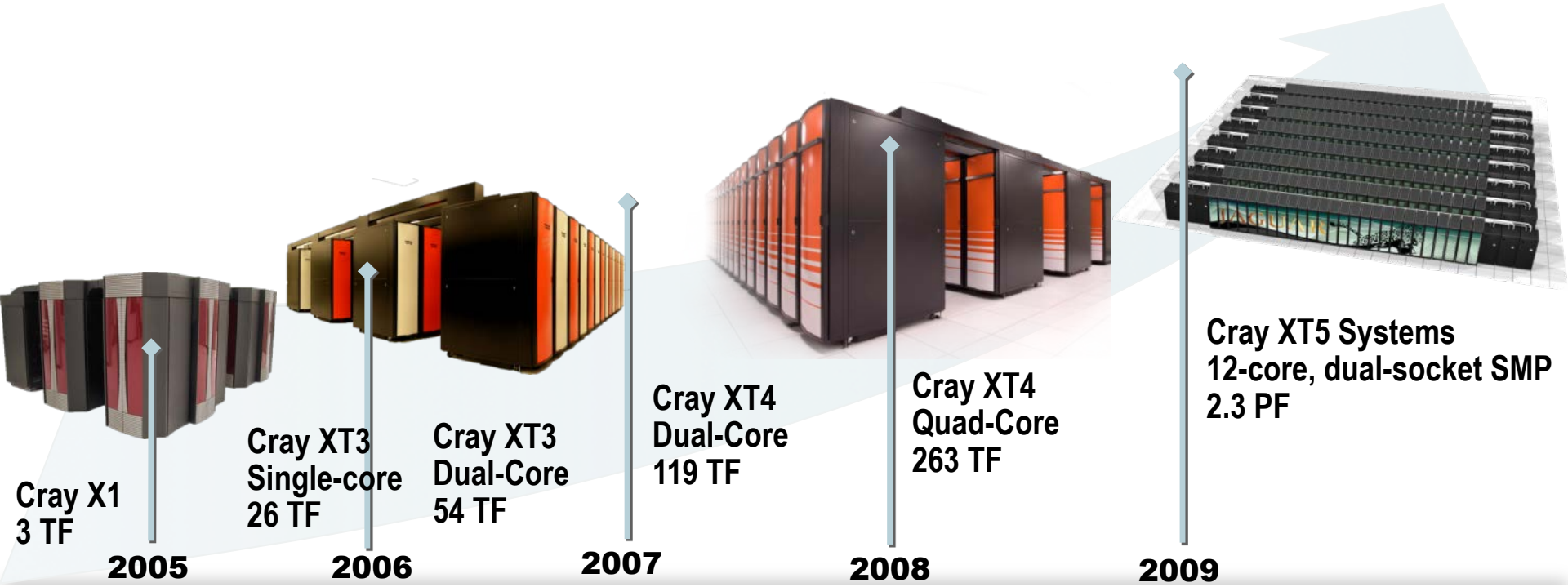
OLCF's Industry Partnership Program has realized sustained, steady growth



- Twenty-six projects are operational in CY 2013, to date.
- Ten percent of Titan's CY 2013 utilization is from industry-led projects

ORNL has increased system performance by 1,000 times 2004-2010

Hardware scaled from single-core through dual-core to quad-core and dual-socket , 12-core SMP nodes	Scaling applications and system software was the biggest challenge



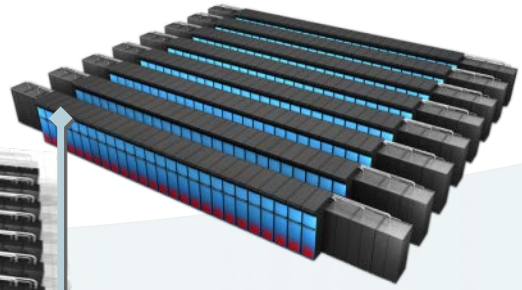
Our Science requires that we advance computational capability 1000x over the next decade.

Mission: Providing world-class computational resources and specialized services for the most computationally intensive global challenges	Vision: Deliver transforming discoveries in climate, materials, biology, energy technologies, etc

Roadmap to Exascale

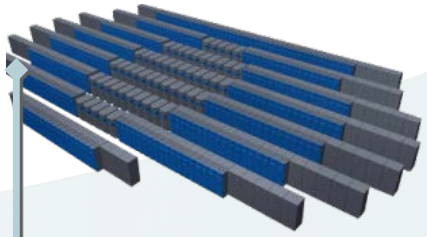
Jaguar 2.3 PF
362 TB DRAM

2010



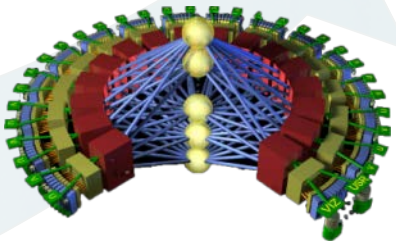
Titan 27 PF
600 TB DRAM
Hybrid GPU/CPU

2012



OLCF-4: 100-250 PF
4000 TB memory
> 20MW
6 day resilience

2017



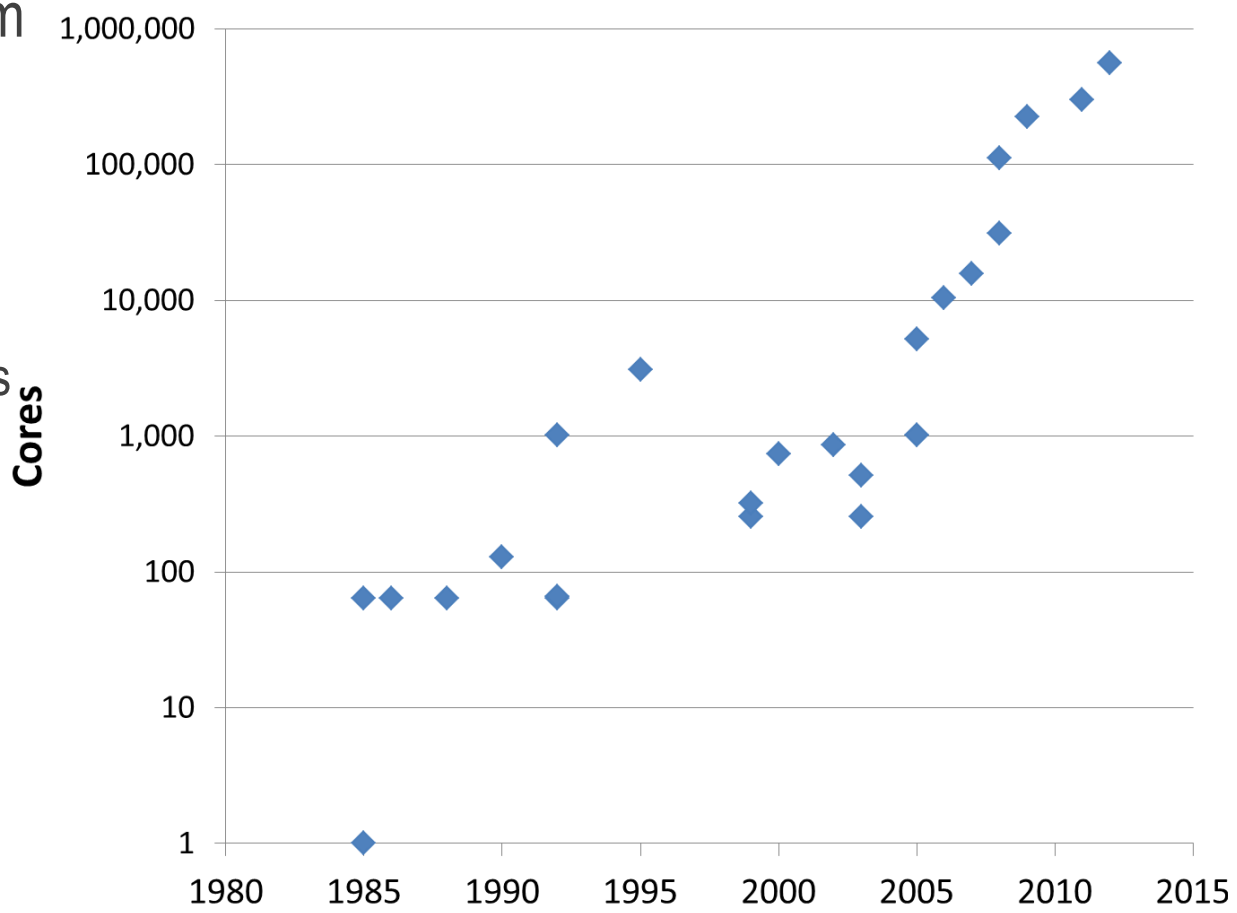
OLCF-5: 1 EF
20 MW

What are the Challenges?

2022

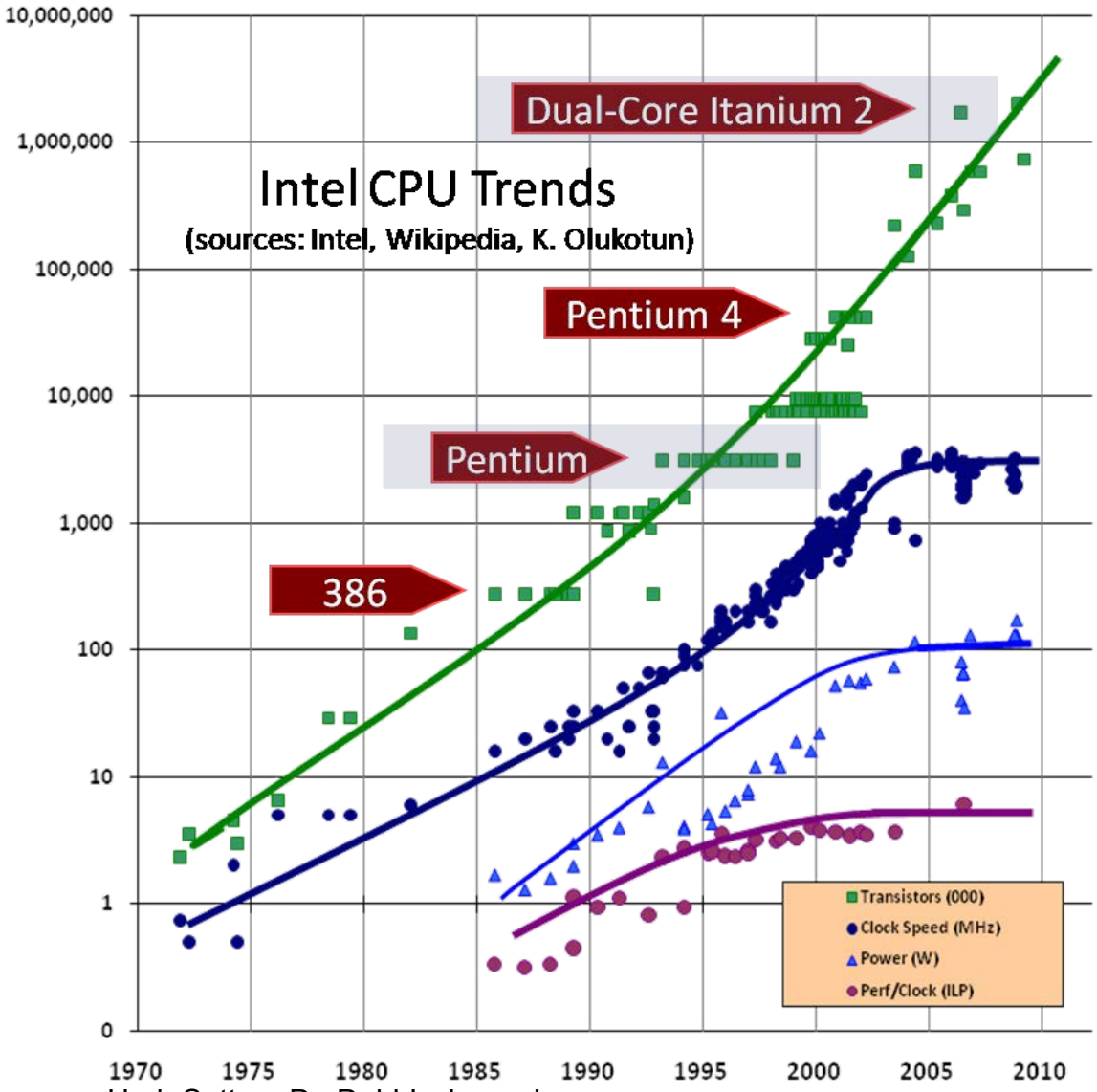
Hardware Trend of ORNL's Systems 1985 - 2013

- In the last 28 years, our systems have scaled from 64 cores to hundreds of thousands of cores and millions of simultaneous threads of execution
  - Multiple hierarchical levels of parallelism
  - Hybrid processors and systems
- The last 28 years of application development have been about finding ways to exploit that parallelism!



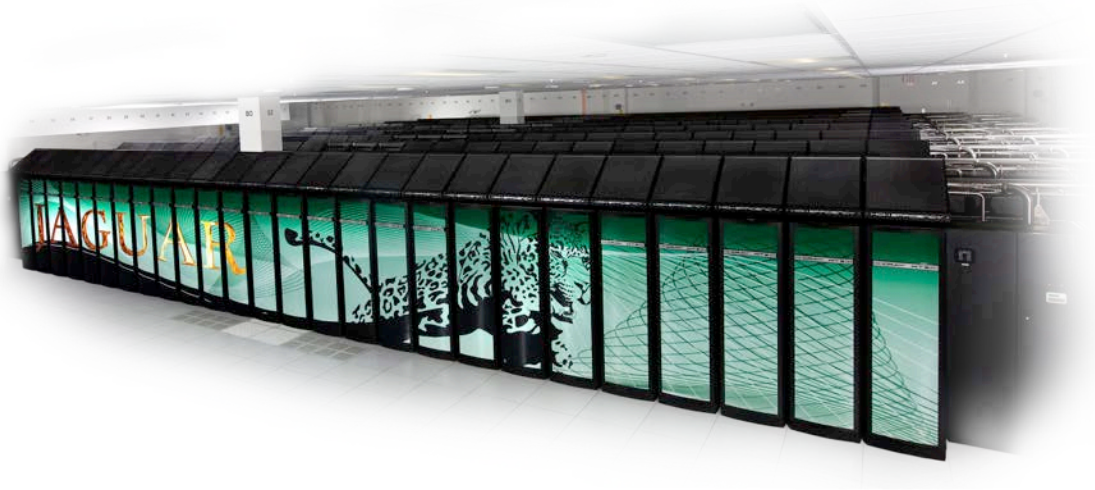
Architectural Trends – No more free lunch

- Moore's Law continues (green line)
- But CPU clock rates stopped increasing in 2003 (dark blue line)
- Power (light blue line) is capped by heat dissipation and $$$
- Single-thread performance is growing slowly (magenta line)



Herb Sutter: Dr. Dobb's Journal:
http://www.gotw.ca/publications/concurrency-ddj.htm

Power is THE problem



Power consumption of 2.3 PF (Peak) Jaguar:
7 megawatts, equivalent to that of a small city (5,000 homes)

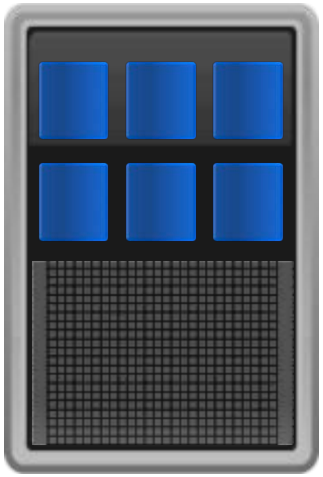
Using traditional CPUs is not economically feasible

20 PF+ system:
30 megawatts (30,000 homes)

Why GPUs? Heirarchical Parallelism
High performance and power efficiency on path to exascale

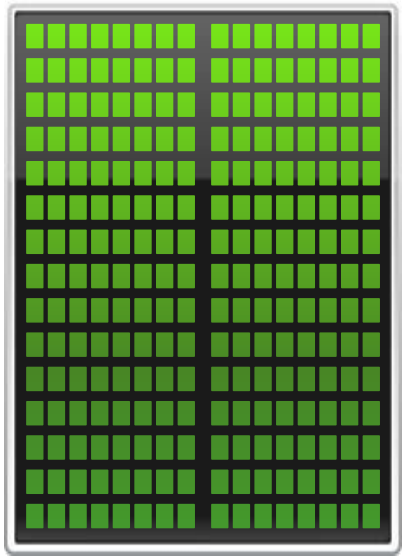
- Hierarchical parallelism improves scalability of applications
- Expose more parallelism through code refactoring and source code directives
  - Doubles performance of many codes
- Heterogeneous multicore processor architecture: Using right type of processor for each task
- Data locality: Keep data near processing
  - GPU has high bandwidth to local memory for rapid access
  - GPU has large internal cache
- Explicit data management: Explicitly manage data movement between CPU and GPU memories

CPU



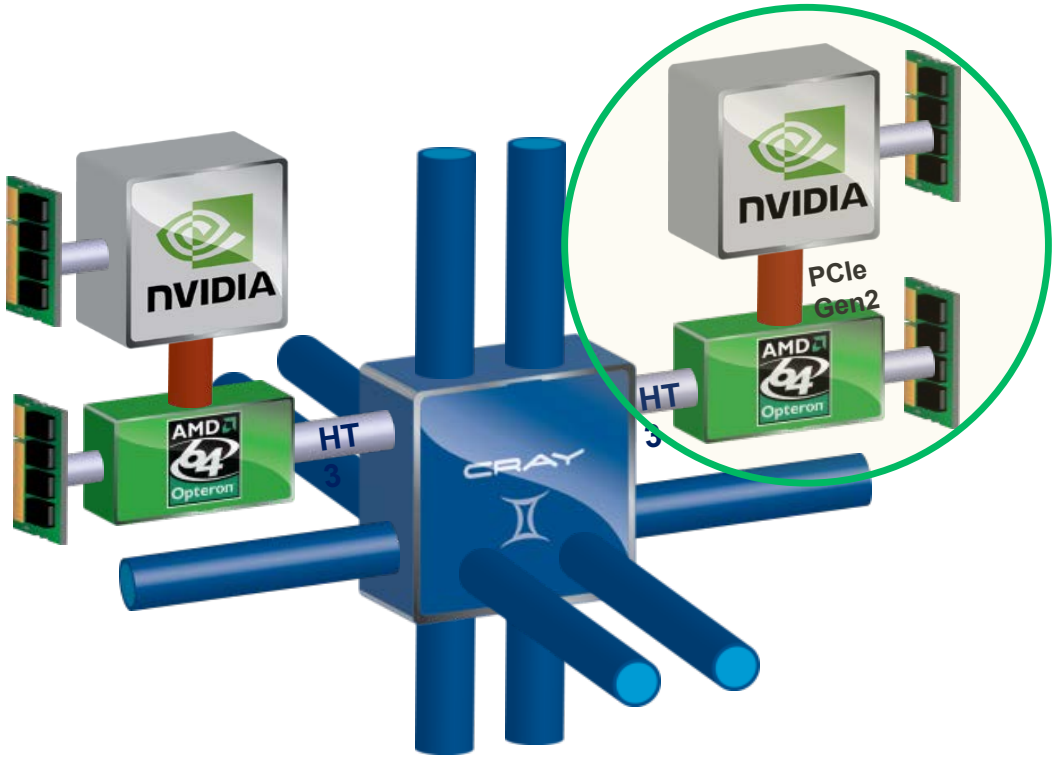
- Optimized for sequential multitasking

GPU Accelerator



- Optimized for many simultaneous tasks
- 10× performance per socket
- 5× more energy-efficient systems

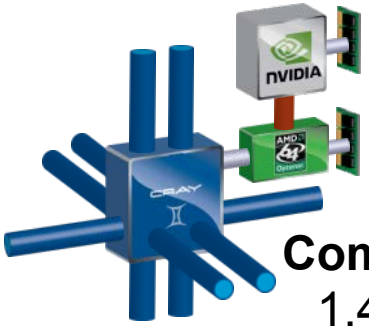
Titan Compute Nodes (Cray XK7)			
Node	AMD Opteron 6200 Interlagos (16 cores)	2.2 GHz	32 GB (DDR3)
Accelerator	Tesla K20x (2688 CUDA cores)	732 MHz	6 GB (DDR5)



Titan:
Cray XK7 System

System:
200 Cabinets
18,688 Nodes
27 PF
710 TB

Cabinet:
24 Boards
96 Nodes
139 TF
3.6 TB



Compute Node:
1.45 TF
38 GB

Board:
4 Compute Nodes
5.8 TF
152 GB

	Titan System (Cray XK7)		
Peak Performance	27.1 PF 18,688 compute nodes	24.5 PF GPU	2.6 PF CPU
System memory	710 TB total memory		
Interconnect	Gemini High Speed Interconnect	3D Torus	
Storage	Luster Filesystem	32 PB	
Archive	High-Performance Storage System (HPSS)	29 PB	
I/O Nodes	512 Service and I/O nodes		

#2

TOP500 SUPERCOMPUTER SITES

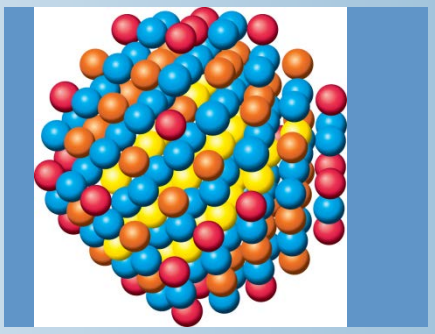
8.2 Megawatts
27 Pflops (Peak)
17.59 PFlops (Linpack)

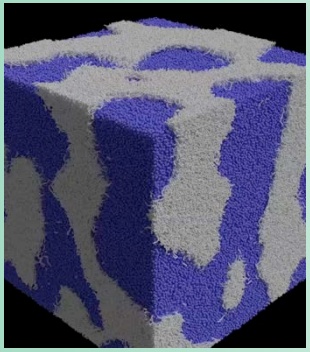
Center for Accelerated Application Readiness (CAAR)

- We created CAAR as part of the Titan project to help prepare applications for accelerated architectures
- Goals:
  - Work with code teams to develop and implement strategies for exposing hierarchical parallelism for our users applications
  - Maintain code portability across modern architectures
  - Learn from and share our results
- We selected six applications from across different science domains and algorithmic motifs

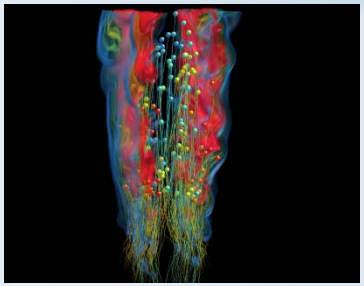
Early Science Challenges for Titan

WL-LSMS
Illuminating the role of material disorder, statistics, and fluctuations in nanoscale materials and systems.





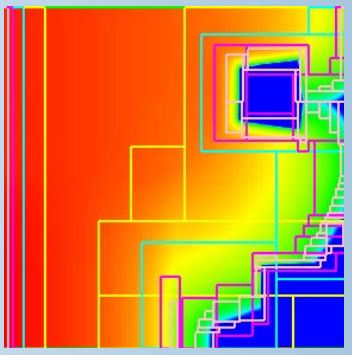
LAMMPS
A molecular dynamics simulation of organic polymers for applications in organic photovoltaic heterojunctions , de-wetting phenomena and biosensor applications

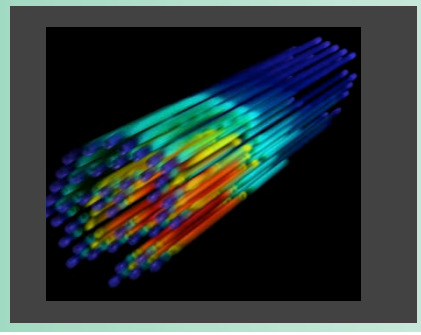


S3D
Understanding turbulent combustion through direct numerical simulation with complex chemistry.
.

CAM-SE
Answering questions about specific climate change adaptation and mitigation scenarios; realistically represent features like precipitation patterns / statistics and tropical storms.

NRDF
Radiation transport – important in astrophysics, laser fusion, combustion, atmospheric dynamics, and medical imaging – computed on AMR grids.





Denovo
Discrete ordinates radiation transport calculations that can be used in a variety of nuclear energy and technology applications.

CAAR Plan

- **Comprehensive team assigned to each app**
  - OLCF application lead
  - Cray engineer
  - NVIDIA developer
  - Other: other application developers, local tool/library developers, computational scientists
- **Single early-science problem targeted for each app**
  - Success on this problem is ultimate metric for success
- **Particular plan-of-attack different for each app**
  - WL-LSMS – dependent on accelerated ZGEMM
  - CAM-SE– pervasive and widespread custom acceleration required
- **Multiple acceleration methods explored**
  - WL-LSMS – CULA, MAGMA, custom ZGEMM
  - CAM-SE– CUDA, directives
  - Two-fold aim
    - **Maximum acceleration for model problem**
    - **Determination of optimal, reproducible acceleration path for other applications**

Effectiveness of GPU Acceleration?

OLCF-3 Early Science Codes -- Performance on Titan XK7

Application	Cray XK7 vs. Cray XE6 Performance Ratio*
LAMMPS* Molecular dynamics	**7.4**
S3D Turbulent combustion	**2.2**
Denovo 3D neutron transport for nuclear reactors	**3.8**
WL-LSMS Statistical mechanics of magnetic materials	**3.8**

Titan: Cray XK7 (Kepler GPU plus AMD 16-core Opteron CPU)
Cray XE6: (2x AMD 16-core Opteron CPUs)
*Performance depends strongly on specific problem size chosen

Additional Applications from Community Efforts

Current Performance Measurements on Titan

Application	Cray XK7 vs. Cray XE6 Performance Ratio*
AWP-ODC Seismology	**2.1**
DCA++ Condensed Matter Physics	**4.4**
QMCPACK Electronic structure	**2.0**
RMG (DFT – real-space, multigrid) Electronic Structure	**2.0**
XGC1 Plasma Physics for Fusion Energy R&D	**1.8**

Titan: Cray XK7 (Kepler GPU plus AMD 16-core Opteron CPU)
Cray XE6: (2x AMD 16-core Opteron CPUs)
*Performance depends strongly on specific problem size chosen

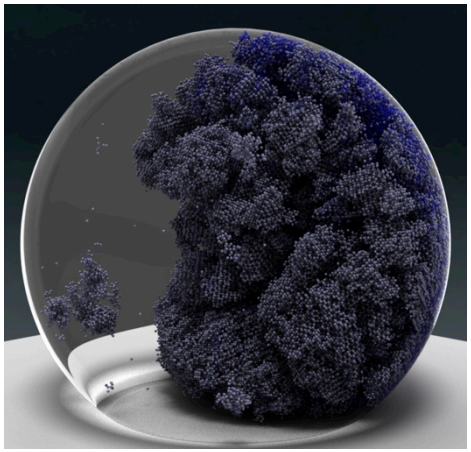
Non-Icing Surfaces for Cold Climate Wind Turbines

Molecular Dynamics Simulations

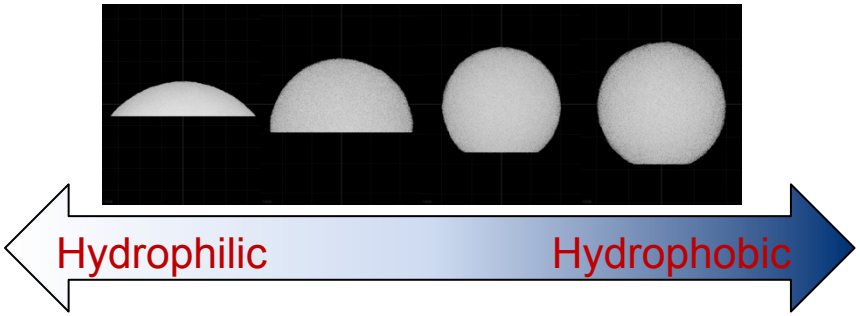
ALCC Program
Masako Yamada
GE Global Research
40 M Titan core hours

Science Objectives and Impact

- **Driver:** Understand microscopic mechanism of critical nucleus formation at water droplet/substrate interface
- **Strategy:** Determine efficacy of non-icing surfaces at different operation temperatures
- **Impact:** Ice accumulation on wind turbines affects efficiency of energy generation, leading to turbine downtime. Potential losses can range from around 10-20%, with an upwards of 50% in the harshest environments



Location of ice nucleation varies dependent on temperature and contact angles. Visualization by M. Matheson (ORNL)



Performance Achievements

- Achieved factor 5X speed-up from GPU acceleration
- Recast three-body, Stillinger-Weber potential to run effectively on the GPU/CPU hybrid Titan (Mike Brown, ORNL)
- Achieved factor 40X speed-up from new mW interaction potential for water, rather than SPC/E:

Science Results

Replicated GE's experimental results:

- High-contact-angle surfaces delay the onset of nucleation
- The delay becomes less pronounced at lower temperatures

Competitiveness Impact

- Improve adoption of wind energy in cold climate regimes by reducing turbine downtime
- Shift from energy intensive active ice-mitigation strategies (which can require up to 25% of the total rated output) to passive non-icing surfaces

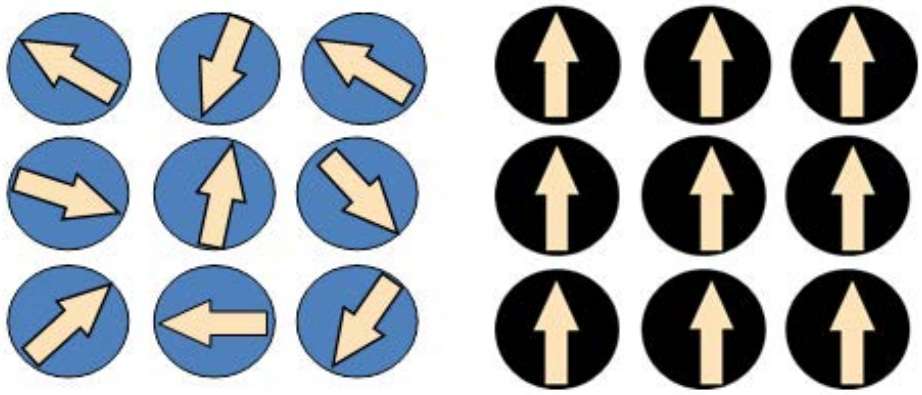
Magnetic Materials

Simulating nickel atoms pushes double-digit petaflops

WL-LSMS
Marcus Eisenbach,
ORNL

Science Objectives and Impact

- Simulate the atomic magnetic direction and strength of nickel
- Enhance the understanding of microscopic behavior of magnetic materials
- Enable the simulation of new magnetic materials
  - Better, cheaper, more abundant materials
- Model development on Titan will enable investigation on smaller computers



Researchers using Titan are studying the behavior of magnetic systems by simulating nickel atoms as they reach their Curie temperature—the threshold between order (right) and disorder (left) when atoms spin into random magnetic directions of fluctuating magnetic strengths, causing the material to lose its magnetism.

Titan Simulation: WL-LSMS

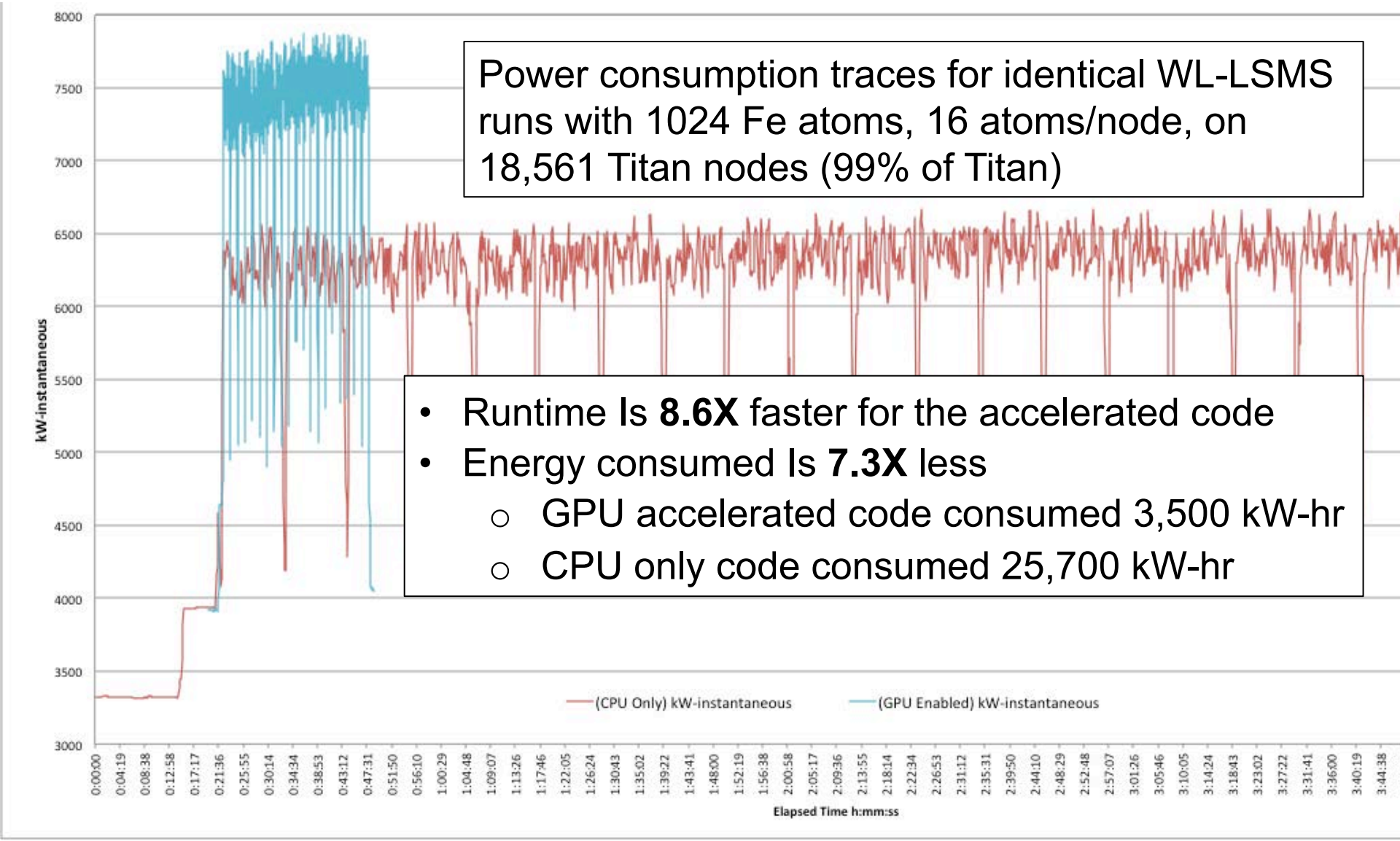
- Combination of first-principles code for magnetic materials and Wang-Landau algorithm for statistical behavior
- More than an 8-factor speedup on Titan compared to Jaguar, Cray XT-5
  - From 1.84 PF to 14.5 PF
- Wang-Landau allows for calculations at realistic temperatures

Preliminary Science Results

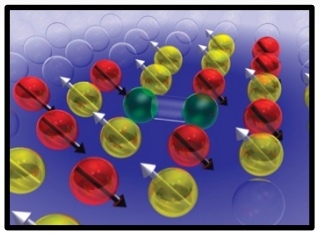
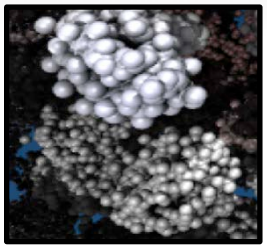
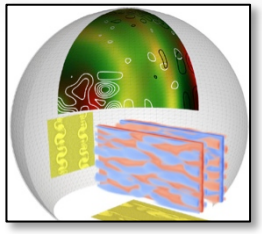
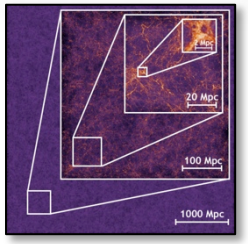
- Titan necessary to calculate nickel's Curie temperature, a more complex calculation than iron
- Calculated 50 percent larger phase space
- Four times faster on Titan than on comparable CPU-only system, (i.e., Cray XE6).

Application Power Efficiency of the Cray XK7

WL-LSMS for CPU-only and Accelerated Computing



Four of Six SC13 Gordon Bell Finalists Used Titan

Peter Staar ETH Zurich	Massimo Bernaschi ICNR-IAC Rome	Michael Bussmann HZDR - Dresden	Salman Habib Argonne
High-Temperature Superconductivity	**Biofluidic Systems**	**Plasma Physics**	**Cosmology**
Taking a Quantum Leap in Time to Solution for Simulations of High-T_C Superconductors	20 Petaflops Simulation of Protein Suspensions in Crowding Conditions	Radiative Signatures of the Relativistic Kelvin-Helmholtz Instability	HACC: Extreme Scaling and Performance Across Diverse Architectures
Titan (13.6 PF)	Titan (19.8 PF)	Titan (7.2 PF)	Sequoia (13.9 PF), Titan
			

Hybrid Programming Model

- On Jaguar, with 299,008 cores, we were seeing scaling issues for single-level, MPI-only applications
- To take advantage of the vastly larger parallelism in Titan, users need to use hierarchical parallelism in their codes
  - Distributed memory: MPI, SHMEM, PGAS
  - Node Local: OpenMP, Pthreads, local MPI communicators
  - Within threads: Vector constructs on GPU, libraries, OpenACC
- ***These are the same types of constructs needed on all multi-PFLOPS computers to scale to the full size of the systems!***

All Codes Will Need Rework To Scale!

- **Up to 1-2 person-years required to port each code from Jaguar to Titan**
  - Takes work, but an unavoidable step **required for exascale regardless of the type of processors.** It comes from the required level of **parallelism on the node**
  - Also **pays off for other systems**—the ported codes often run significantly faster CPU-only (Denovo 2X, CAM-SE >1.7X)
- We estimate possibly **70-80% of developer time is spent in code restructuring**, regardless of whether using OpenMP / CUDA / OpenCL / OpenACC / …
- **Each code team must make its own choice of using OpenMP vs. CUDA vs. OpenCL vs. OpenACC**, based on the specific case—may be different conclusion for each code
- **Our users and their sponsors must plan for this expense.**

Rethink your algorithms

- **Heterogeneous architectures can make previously infeasible or inefficient models and implementations viable**
  - Alternative methods for electrostatics that perform slower on traditional x86 can be significantly faster on GPUs *(Nguyen, et al. J. Chem. Theor. Comput. 2013. 73-83)*
  - 3-body coarse-grain simulations of water with greater concurrency can allow > 100X simulation rates when compared to fastest atomistic models even though both are run on the GPUs *(Brown, et al. Submitted)*

Programming Environment Components
(Having multiple options improve performance opportunities and reduce risk)

- Compilers
  - Cray Compiler Environment – add GPU support
  - HMPP from CAPS – add C++, Fortran, and additional GPU support
  - PGI – new GPU support
- Debuggers
  - DDT – Add GPU support
  - Scalability already being addressed outside the project
- Performance Analysis tools
  - Cray – add GPU support
  - Vampir suite – add GPU support, and increase scalability
  - Supply other third party tools (HPCToolkit, TAU)
- Math libraries
  - Cray – add GPU support
  - Third party – CULA, MAGMA

Points to ponder

- Science codes are under active development—porting to GPU can be pursuing a “moving target,” challenging to manage
- More available FLOPS on the node should lead us to think of new science opportunities enabled—e.g., more degrees of freedom per grid cell
- We may need to look in unconventional places to get another ~30X thread parallelism that may be needed for exascale—e.g., parallelism in time

OLCF User Requirements Survey – Key Findings

- Surveys are a "lagging indicator" that tend to tell us what problems the users are seeing now, not what they expect to see in the future
- *For the first time ever, FLOPS was NOT the #1 ranked requirement*
- Local memory capacity was not a driver for most users, perhaps in recognition of cost trends
- 76% of users said there is still a moderate to large amount of parallelism to extract in their code, but…
- 85% of respondents rated the difficulty level of extracting that parallelism as moderate or difficult – underlining the need for more training and assistance in this area
- Data sharing, long-term aggregate archival storage of 100s of PB, data analytics, and faster WAN connections also showed up as important requirements

Hardware feature	Ranking
Memory Bandwidth	4.4
Flops	4.0
Interconnect Bandwidth	3.9
Archival Storage Capacity	3.8
Interconnect Latency	3.7
Disk Bandwidth	3.7
WAN Network Bandwidth	3.7
Memory Latency	3.5
Local Storage Capacity	3.5
Memory Capacity	3.2
Mean Time to Interrupt	3.0
Disk Latency	2.9

Rankings from OLCF users
1=not important, 5=very important

Are we seeing the real challenges?

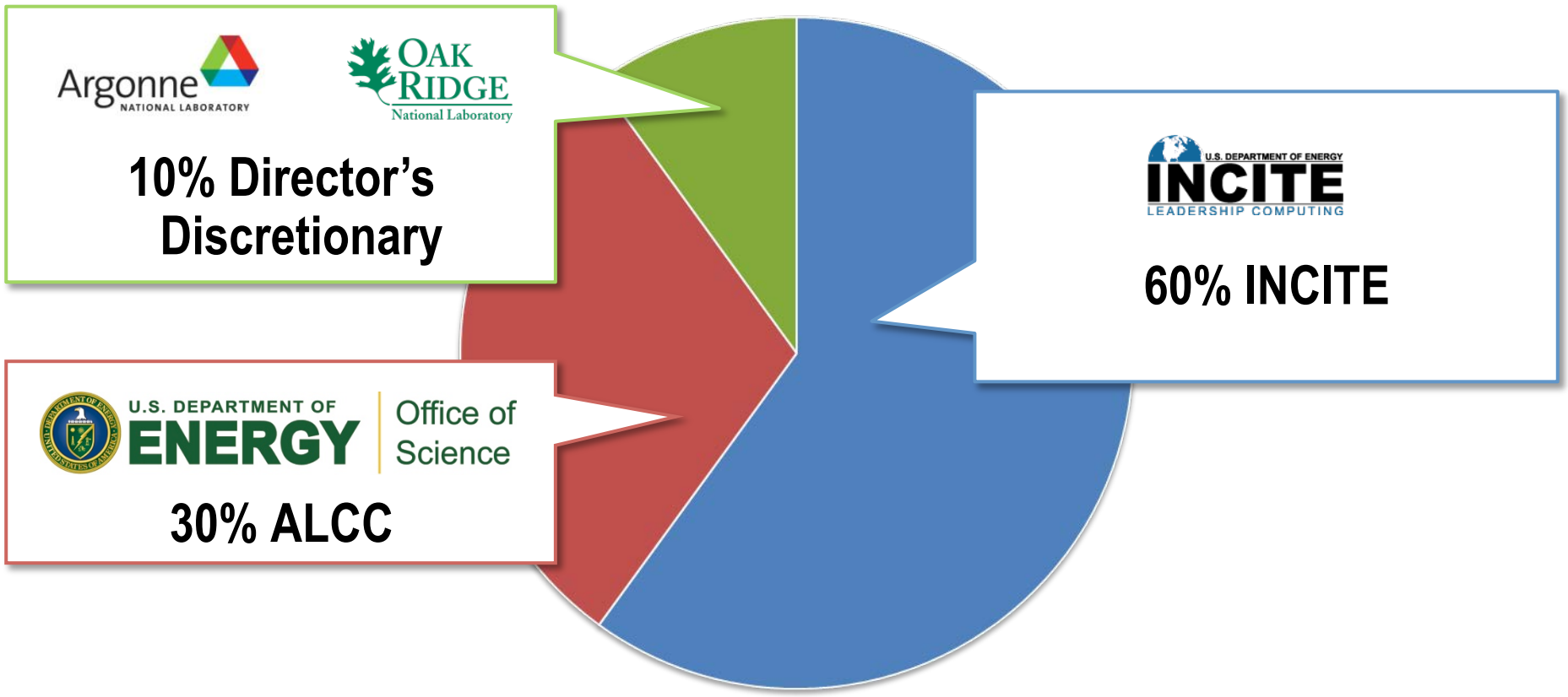
Advanced simulation and modeling apps

Conquering Petascale problems of today

Beware being eaten alive by the Exascale problems of tomorrow.

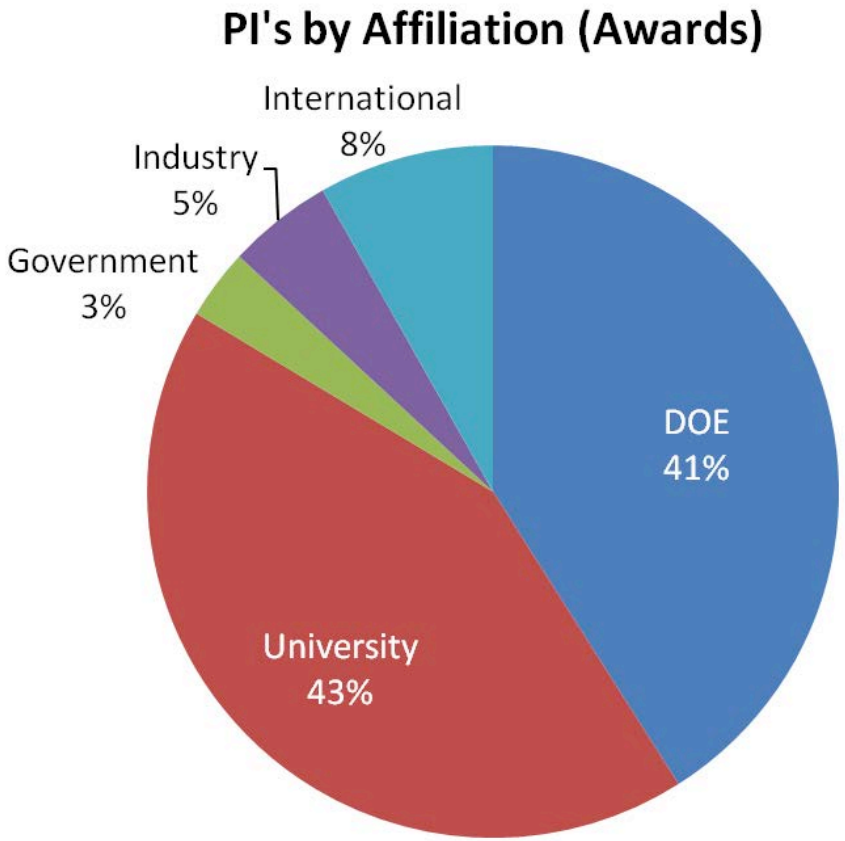
Scale
Power
Resilience
Memory wall

LCF User Programs:
More than 6.5 billion core hours awarded in 2013



User programs are open and competitive

- Total INCITE requests **~15 billion core-hours**, 3x more than the 5 billion core-hours requested last year
- Number of proposals submitted increased **nearly 20%**
- Awards of ~5 billion core-hours for CY 2013
- **61 projects awarded of which 20 are renewals**



Acceptance rates

33% of nonrenewal submittals and 100% of renewals

Contact information
Julia C. White, INCITE Manager
whitejc@DOEleadershipcomputing.org

Getting Started at OLCF: *Project Allocation Requests*

	INCITE	ALCC	Director's Discretionary (DD)
Allocations	Large (Avg. 70 M core hours)	Large (Avg. 40 M core hours)	Medium (3M core hours)
Call for Proposals	Once per year (April open, June close)	Once per year (October open, February close)	Continuous
Projects start	January	July	Rolling
Duration	1-3 years	1 year	1 year
Priority	High	High	Medium

https://www.olcf.ornl.gov/support/getting-started/

Conclusions

- Leadership computing is for the critically important problems that need the most powerful compute and data infrastructure
- Our compute and data resources have grown 10,000X over the decade, are in high demand, and are effectively used.
- Computer system performance increases through parallelism
  - Clock speed trend flat to slower over coming years
  - Applications must utilize all inherent parallelism
- Accelerated, hybrid-multicore computing solutions are performing well on real, complex scientific applications.
  - But you must work to expose the parallelism in your codes.
- OLCF resources are available to industry, academia, and labs, through open, peer-reviewed allocation mechanisms.

Acknowledgements

OLCF-3 CAAR Team: Bronson Messer, Wayne Joubert, Mike Brown, Matt Norman, Markus Eisenbach, Ramanan Sankaran

OLCF-3 Vendor Parters: Cray, AMD, and NVIDIA

- This research used resources of the Oak Ridge Leadership Computing Facility at the Oak Ridge National Laboratory, which is supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC05-00OR22725.

Questions?

WellsJC@ornl.gov

Contact us at http://jobs.ornl.gov