Accelerating Science with Titan, the World's Fastest Supercomputer

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

Presented by:

Jack Wells
Oak Ridge Leadership Computing Facility (OLCF)
Oak Ridge National Laboratory (ORNL)

Lattice QCD Workshop
April 29, 2013
Oak Ridge

U.S. DEPARTMENT OF ENERGY

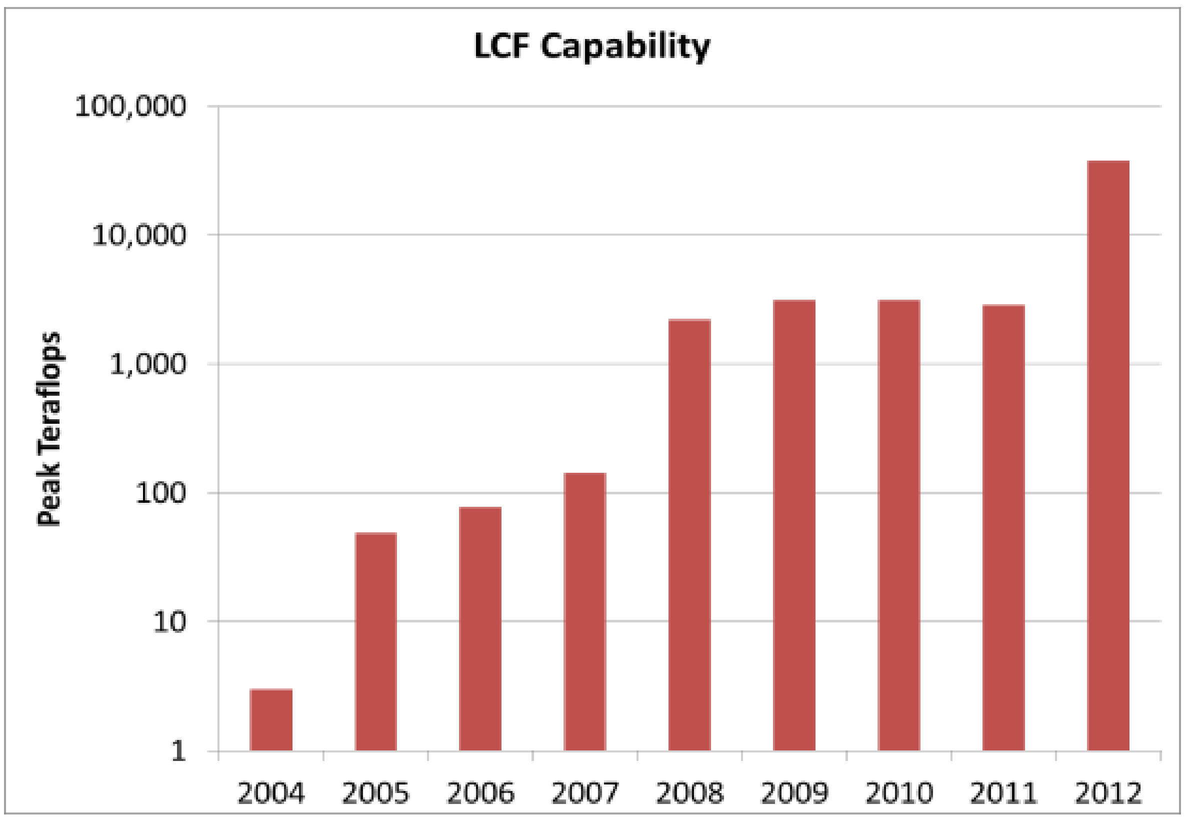
OAK RIDGE NATIONAL LABORATORY
MANAGED BY UT-BATTELLE FOR THE DEPARTMENT OF ENERGY

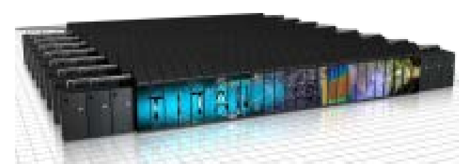
What is the Leadership Computing Facility (LCF)?

- Collaborative DOE Office of Science program at Oak Ridge and Argonne National Laboratories
- Mission: Provide the computational and data science resources required to solve the most important scientific & engineering problems in the world.
- 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).

We have increased our system capability by 10,000 times since our founding in 2004

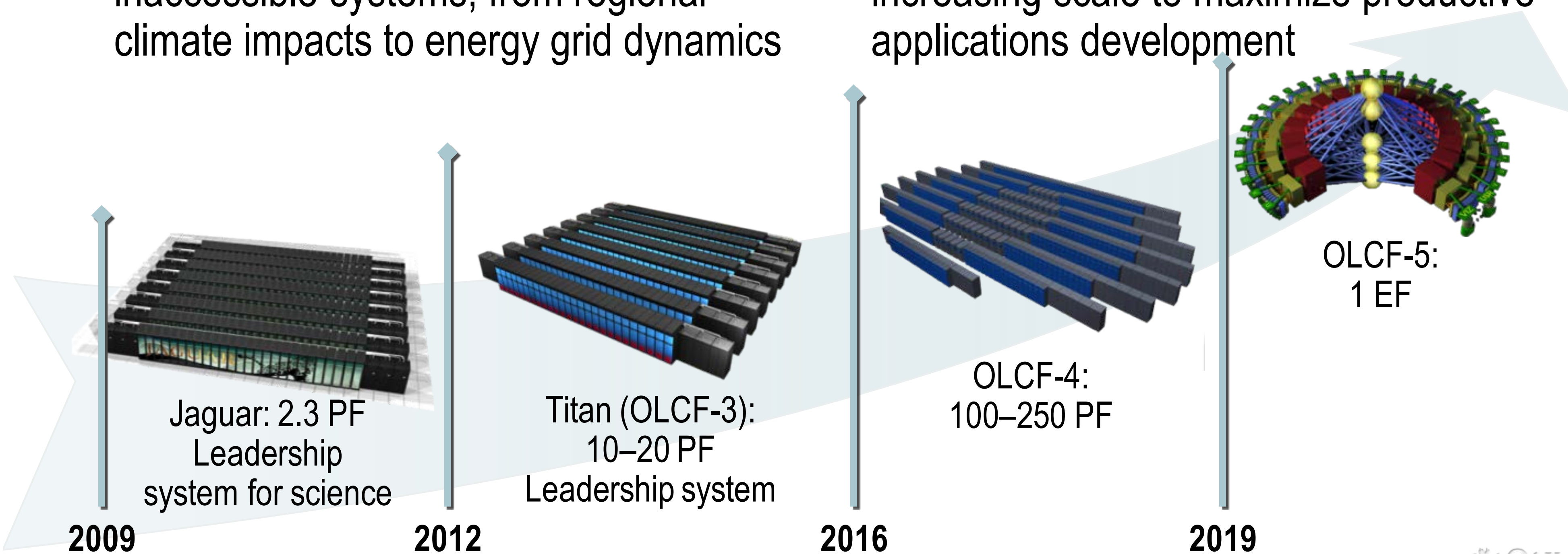
- Strong partnerships with supercomputer vendors.
- LCF users employ large portions of the machine for large fractions of time.
- Strong partnerships with our users to scale codes and algorithms.





Science requires exascale capability in this decade

Mission: Deploy and operate the computational resources required to tackle global challenges	**Vision: Maximize scientific productivity and progress on largest scale computational problems**
• Deliver transforming discoveries in climate, materials, biology, energy technologies, etc. • Enabling investigation of otherwise inaccessible systems, from regional climate impacts to energy grid dynamics	• World-class computational resources and specialized services for the most computationally intensive problems • Stable hardware/software path of increasing scale to maximize productive applications development



Slide courtesy of Jeff Nichols

ASCR Facilities Strategic Planning

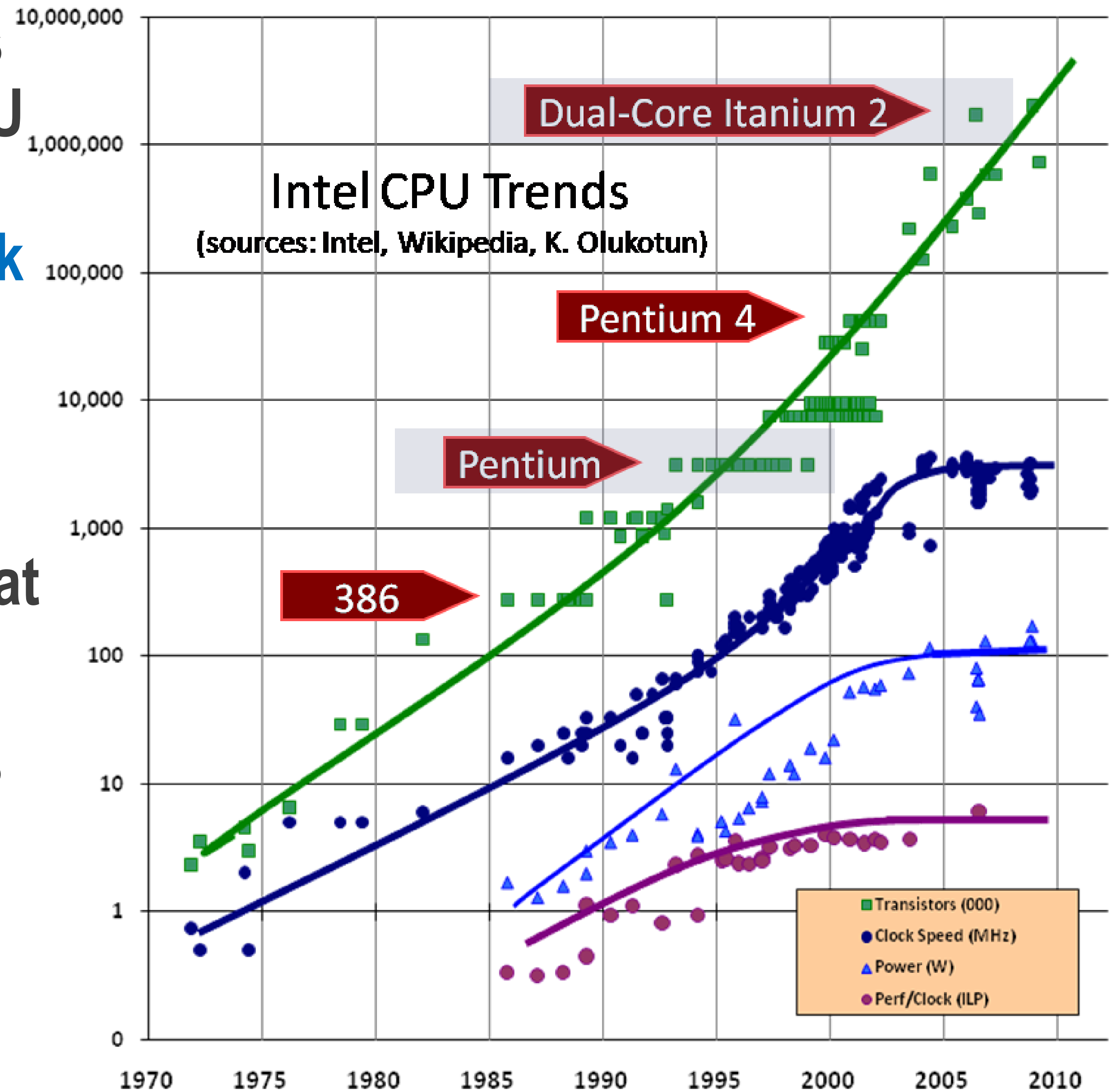
- Objective: To inform the development of an ASCR Facilities 10 year plan
  - Understand and shape the focus of each facility
  - Unify the overarching strategy across facilities
  - Develop plans of action given possible future dependencies, opportunities, and threats
- Constraint: 30 MW of power
- Facilities presented their plans to each other, HQ and ASCAC Facilities Subcommittee on January 30, 2013
- Findings: Need for coordination with other SC office
  - To prepare applications for future architectures
  - To develop a common strategy for addressing SC data analysis and archival needs from simulations and experiments
- Next step: Generate draft ASCR 10-year facilities plan

Overall Acquisition Strategy for Next Systems: Hundreds of Petaflops and beyond

- Joint RFPs between NNSA and SC laboratories
  - Trinity/NERSC-8 (LANL/SNL/LBNL) - Targeting 2015/2016 deployments
  - CORAL (ORNL/ANL/LLNL) - Targeting later deployments
- Open Solicitations
  - Joint selection teams, RFIs, RFPs, but ultimately separate vendor/lab for actual systems
  - RFPs contain core requirements plus lab specific features
- Non-Recurring Engineering Investment coupled with Acquisitions
  - Opportunity for software or technology investments to provide additional features
  - Opportunity for variations through separate contracts

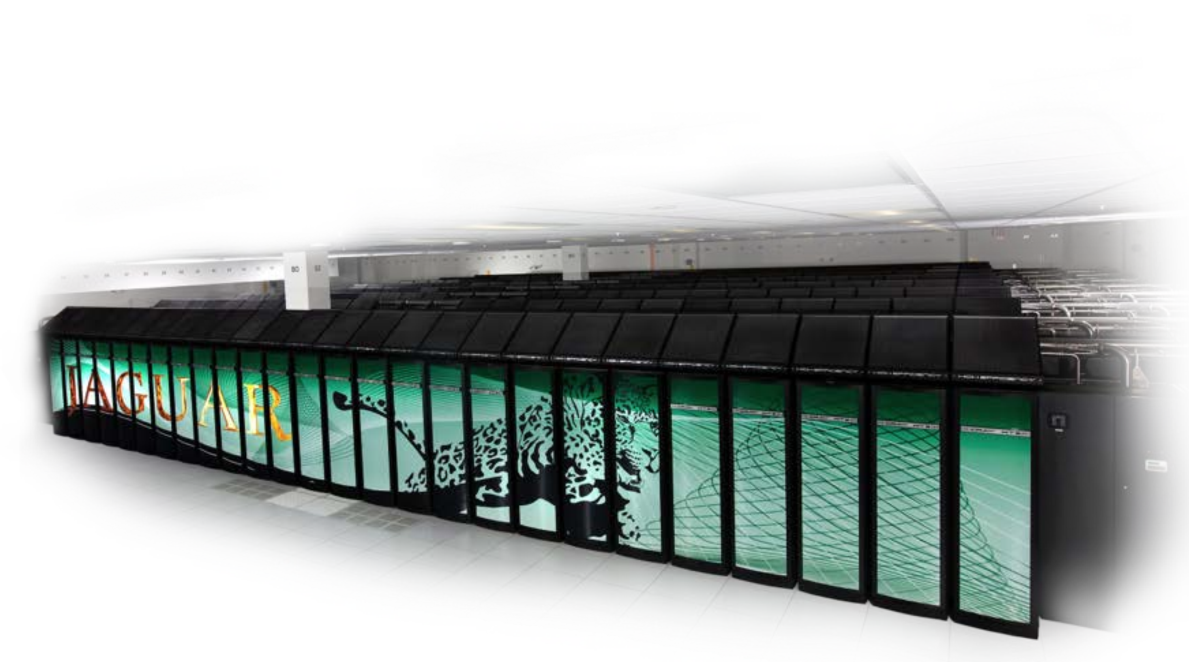
Architectural Trends – No more free lunch

- Moore's Law continues (green line) but the CPU clock rates stopped increasing in 2003 (dark blue line) due to power constraints. (light blue line)
- Power is capped by heat dissipation and $$$
- Performance increases have been coming through increased parallelism



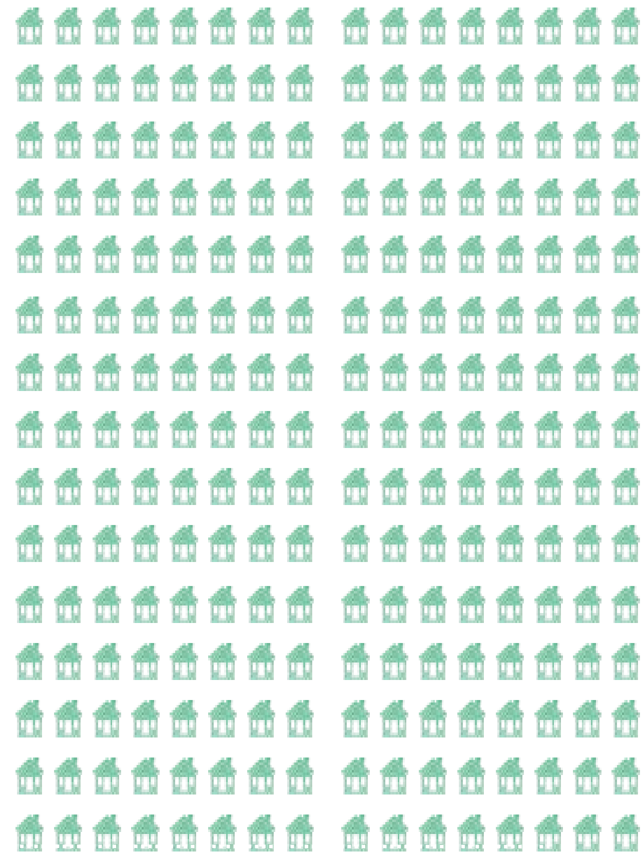
Herb Sutter: Dr. Dobb's Journal:
http://www.gotw.ca/publications/concurrency-ddj.htm

Power is THE problem



Power consumption of 2.3 PF 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)

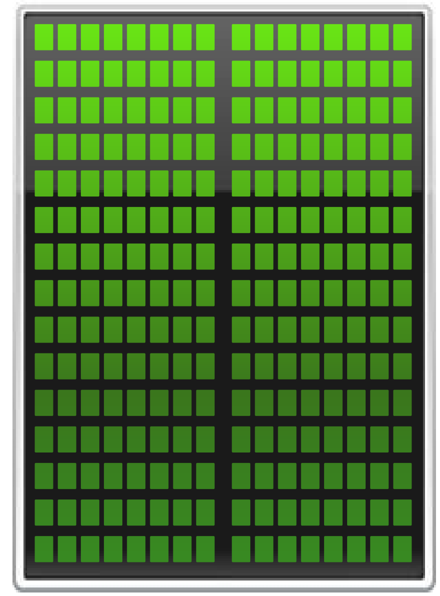
Why GPUs?
High performance and power efficiency on path to exascale

- 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

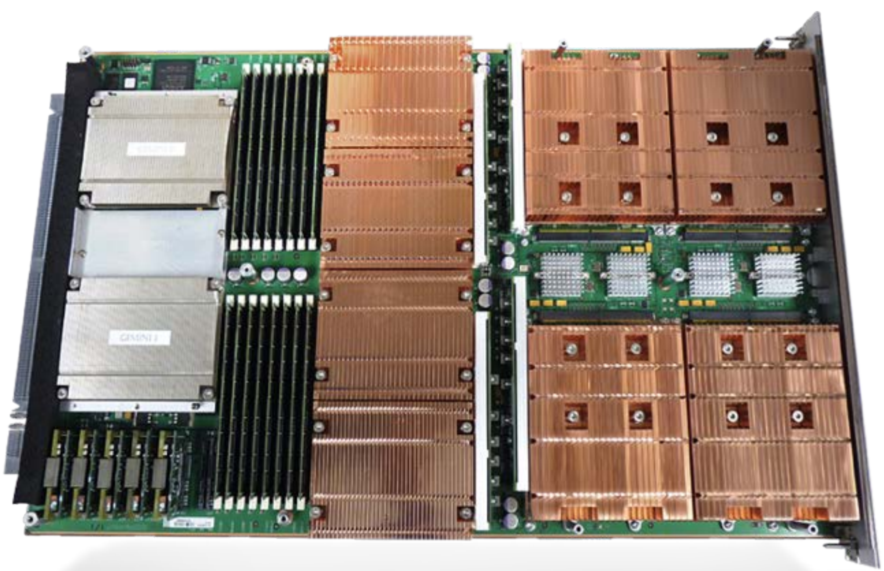
- Optimized for sequential multitasking

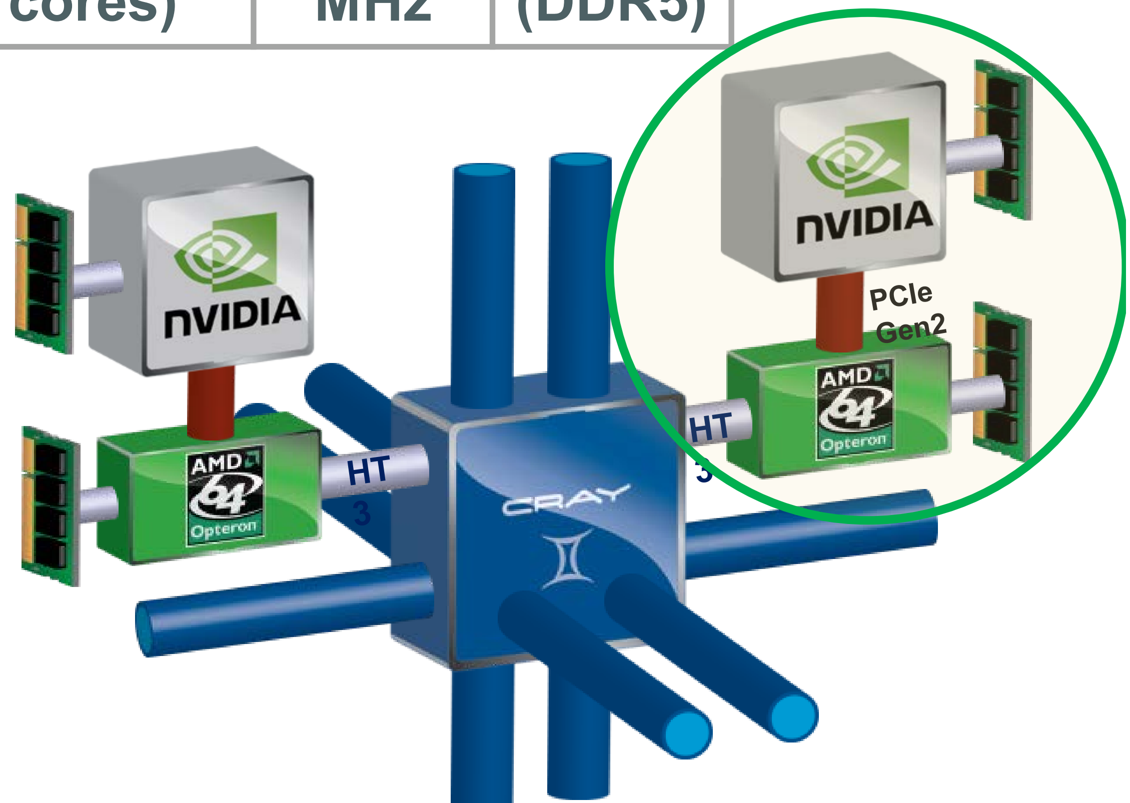
GPU Accelerator



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

	Titan 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)



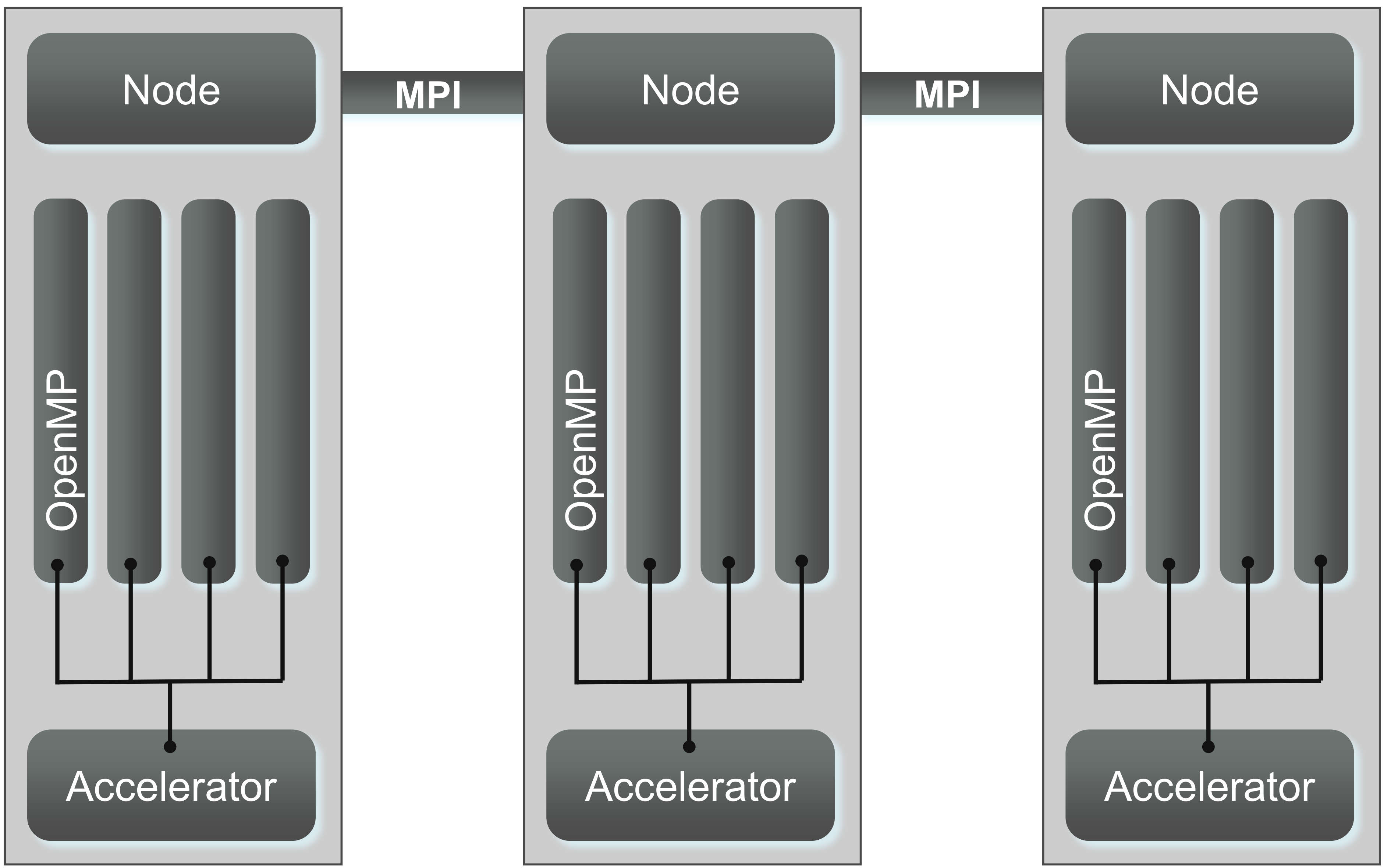


	Titan System (Cray XK7)		
Peak Performance	27.1 PF 18,688 compute nodes	24.5 PF GPU	2.6 PF CPU
System memory	710 TB total memory		
Interconnect	Gemini High Speed Interconnect	3D Torus	
Storage	Luster Filesystem	5 PB	
Archive	High-Performance Storage System (HPSS)	29 PB	
I/O Nodes	512 Service and I/O nodes		

Hybrid Programming Model

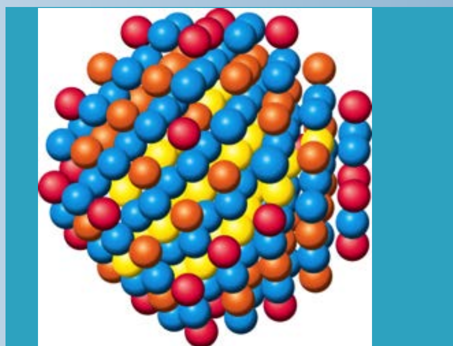
- On Jaguar, with 299,008 cores, we were seeing the limits of a single level of MPI scaling for most applications
- To take advantage of the vastly larger parallelism in Titan, users need to use hierarchical parallelism in their codes
  - Distributed memory: **MPI**, SHMEM, PGAS
  - Node Local: **OpenMP**, 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!***

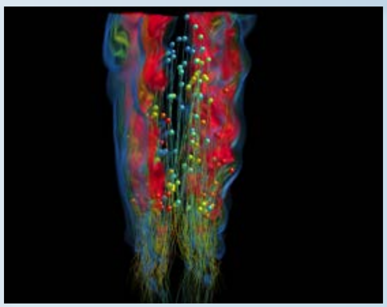
Hybrid Programming Model



Early Science Challenges for Titan

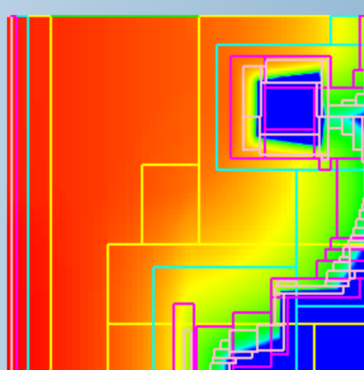
WL-LSMS
Illuminating the role of material disorder, statistics, and fluctuations in nanoscale materials and systems.

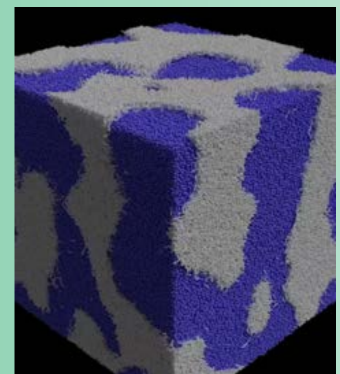




S3D
Understanding turbulent combustion through direct numerical simulation with complex chemistry.

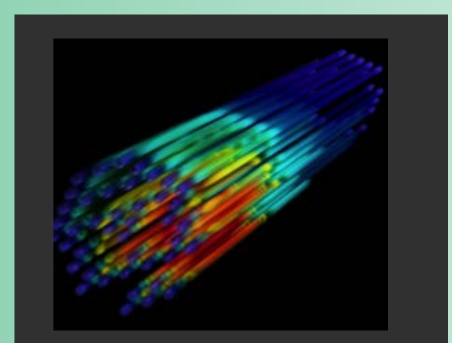
NRDF
Radiation transport – important in astrophysics, laser fusion, combustion, atmospheric dynamics, and medical imaging – computed on AMR grids.





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

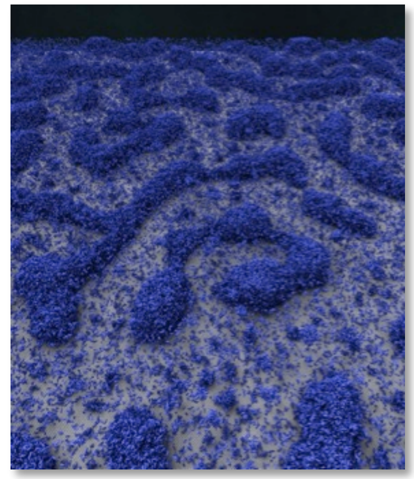
CAM-SE
Answering questions about specific climate change adaptation and mitigation scenarios; realistically represent features like precipitation patterns / statistics and tropical storms.

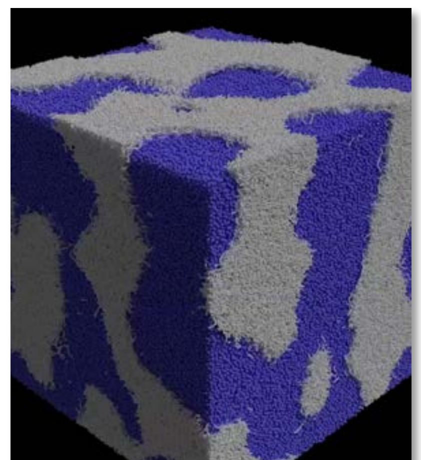


Denovo
Discrete ordinates radiation transport calculations that can be used in a variety of nuclear energy and technology applications.

Leading Molecular Dynamics App Perfect Partner for Titan's GPUs

- LAMMPS is showing excellent results in porting to Titan's GPU-accelerated architecture
- Two areas have had significant success in increased application performance:
  - Organic photovoltaics, or solar cells that use organic molecules instead of traditional semiconductors to convert sunlight into electricity,
  - Liquid crystals that can be used as biomedical sensors to detect bacteria, antibodies, or other signs of illness in the body.
- The increased speed from the GPUs allows Titan to take on even larger systems—many more atoms and molecules — enabling new scientific studies not previously possible.

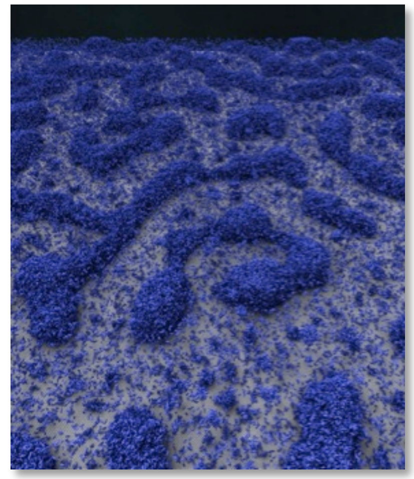

4,900 node Titan simulation of the self-assembly of liquid-crystal molecules on a substrate.


Coarse-grained molecular dynamics simulation of the formation of electron donor and acceptor regions in an organic solar cell device.

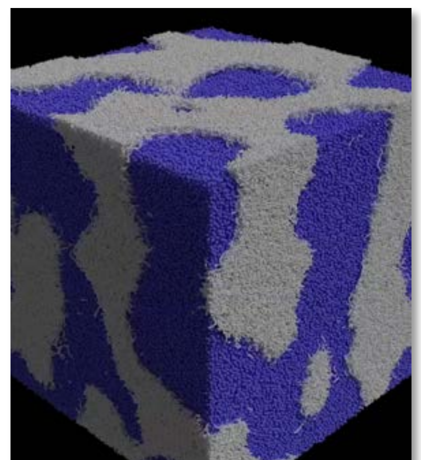
LAMMPS on Titan for liquid-crystal calculations is over 10X faster than Jaguar and 7X faster than Cray's XE6 architecture (dual-CPU/no-GPU per node).

Leading Molecular Dynamics App Perfect Partner for Titan's GPUs

- LAMMPS is showing excellent results in porting to Titan's GPU-accelerated architecture
- Two areas have had significant success in increased application performance:
  - Organic photovoltaics, or solar cells that use organic molecules instead of traditional semiconductors to convert sunlight into electricity,
  - Liquid crystals that can be used as biomedical sensors to detect bacteria, antibodies, or other signs of illness in the body.
- The increased speed from the GPUs allows Titan to take on even larger systems—many more atoms and molecules — enabling new scientific studies not previously possible.



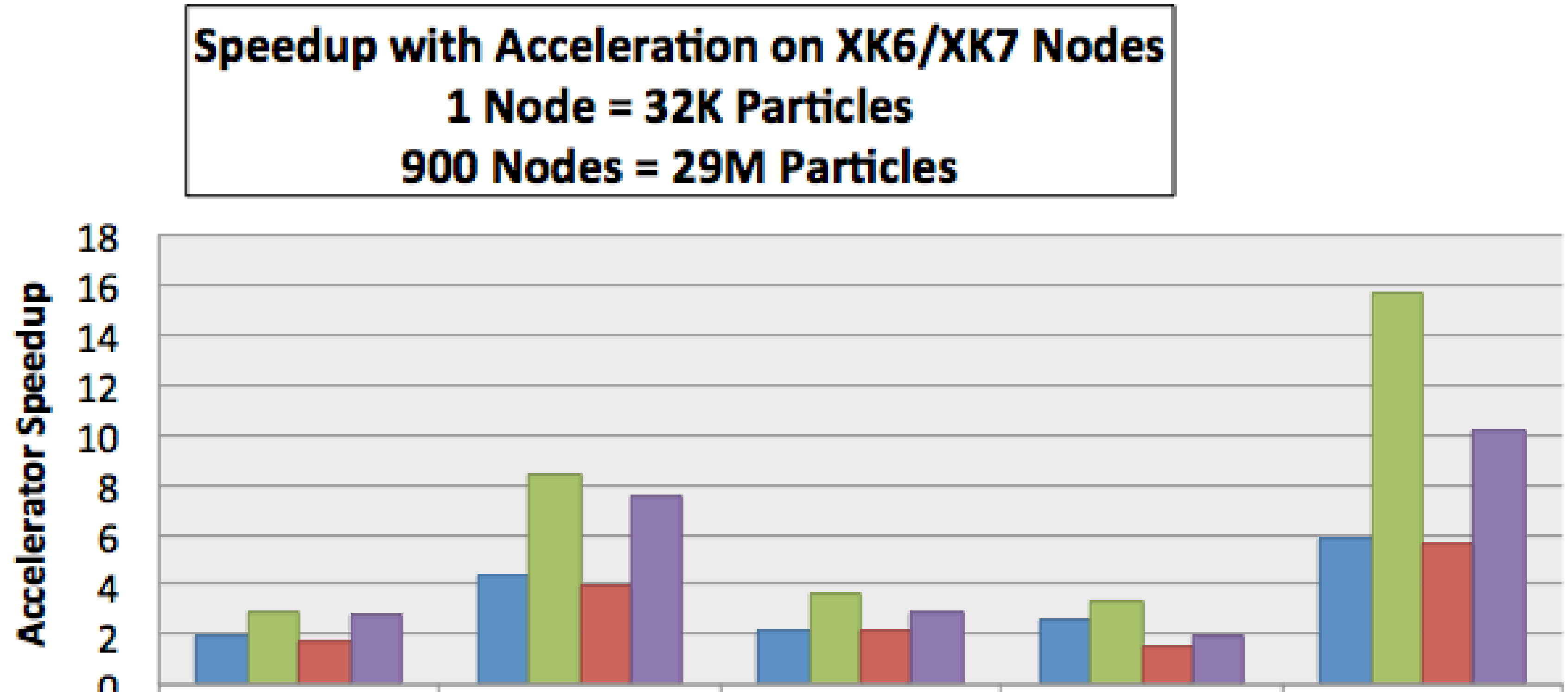
4,900 node Titan simulation of the self-assembly of liquid-crystal molecules on a substrate.



Coarse-grained molecular dynamics simulation of the formation of electron donor and acceptor regions in an organic solar cell device.

LAMMPS on Titan for liquid-crystal calculations is over 10X faster than Jaguar and 7X faster than Cray's XE6 architecture (dual-CPU/no-GPU per node).

LAMMPS Accelerator Speedup



	Atomic Fluid (cutoff = 2.5σ)	Atomic Fluid (cutoff = 5.0σ)	Bulk Copper	Protein	Liquid Crystal
XK6 (1 Node)	1.92	4.33	2.12	2.6	5.82
XK7 (1 Node)	2.90	8.38	3.66	3.36	15.70
XK6 (900 Nodes)	1.68	3.96	2.15	1.56	5.60
XK7 (900 Nodes)	2.75	7.48	2.86	1.95	10.14

Mike Brown (ORNL),
https://www.olcf.ornl.gov/wp-content/uploads/2013/02/Porting-LAMMPS-to-Titan.pdf

How Effective are GPUs on Scalable Applications?

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

Application	Cray XK7 vs. Cray XE6 Performance Ratio*
LAMMPS Molecular dynamics	**7.4**
S3D Turbulent combustion	**2**
Denovo 3D neutron transport for nuclear reactors	**3.8**
WL-LSMS Statistical mechanics of magnetic materials	**3.5**

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 Early Performance Measurements on Titan

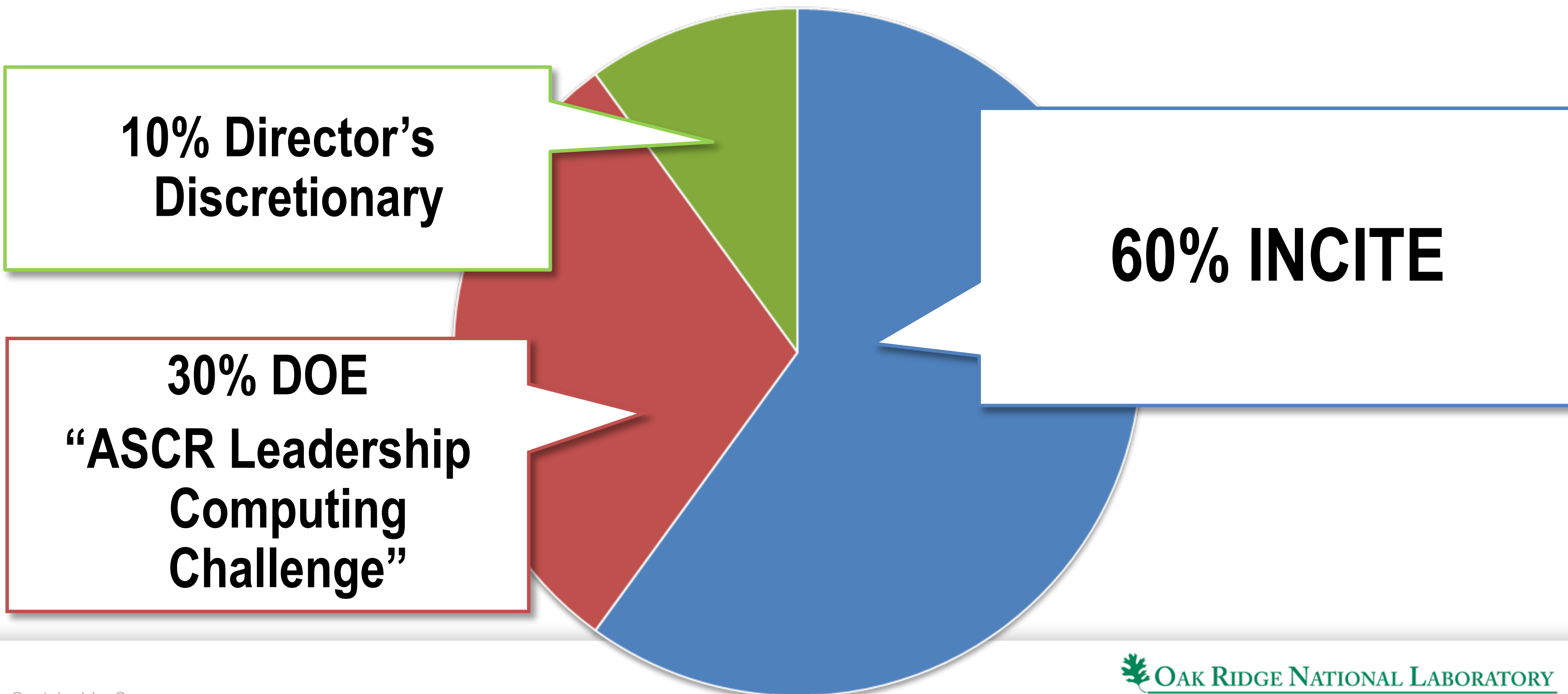
Application	Cray XK7 vs. Cray XE6 Performance Ratio*
AWP-ODC Seismology	**2.1**
DCA++ Condensed Matter Physics	**4.4**
QMCPACK Electronic structure	**2.0**
RMG (DFT – real-space, multigrid) Electronic Structure	**2.0**
XGC1 Plasma Physics for Fusion Energy R&D	**1.8**

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

DOE Computational Facilities Allocation Policy for Leadership Facilities

Primary Objective:

- *"Provide substantial allocations to the open science community through an peered process for a small number of high-impact scientific research projects"*



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.
- OLCF is planning now for ~200PF computing systems in 2016-2017
- 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 very well on real, complex scientific applications.
- OLCF resources are available to academia and industry 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?