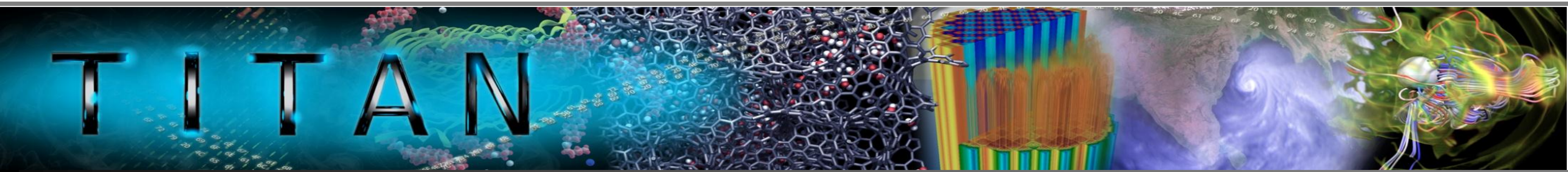
SC12
Salt Lake City, Utah

Titan - Early Experience with the Titan System at Oak Ridge National Laboratory

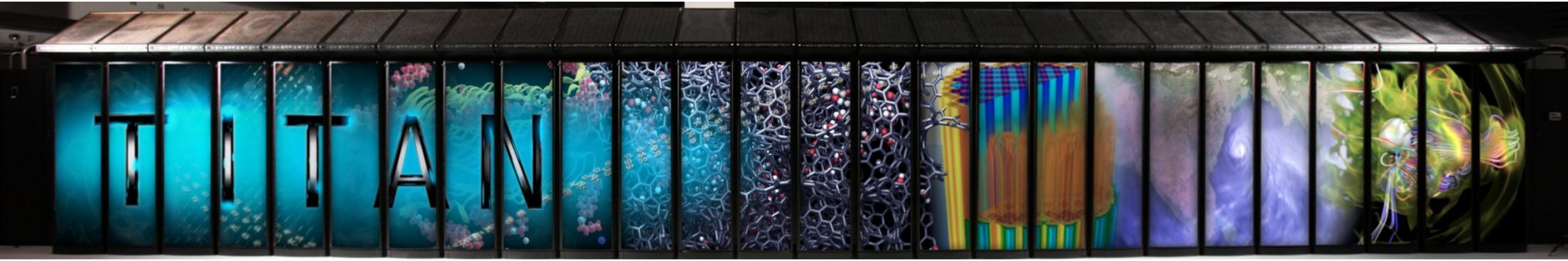
20 Years of Excellence in Computational Science
OLCF
Oak Ridge Leadership Computing Facility
1992–2012

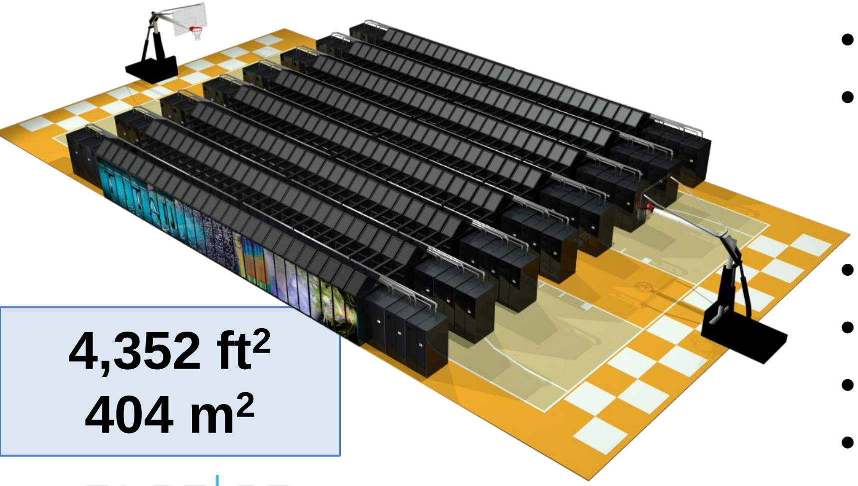
Buddy Bland
Project Director
Oak Ridge Leadership Computing Facility

November 13, 2012



ORNL's "Titan" Hybrid System: Cray XK7 with AMD Opteron and NVIDIA Tesla



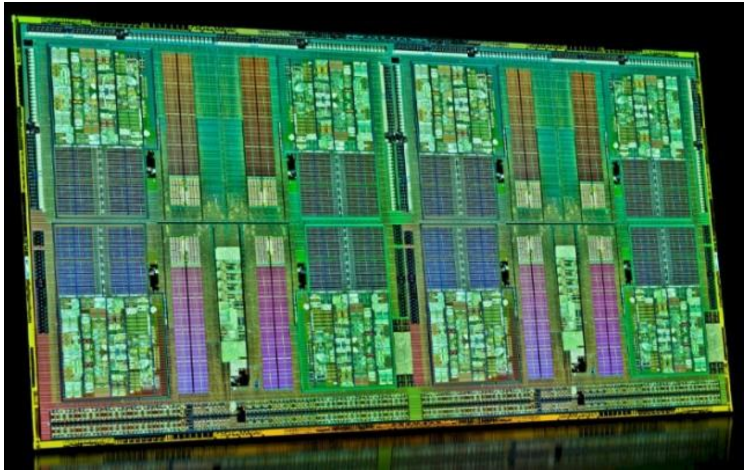


4,352 ft^2
404 m^2

SYSTEM SPECIFICATIONS:

- Peak performance of 27.1 PF (24.5 & 2.6)
- 18,688 Compute Nodes each with:
  - 16-Core AMD Opteron CPU (32 GB)
  - NVIDIA Tesla "K20x" GPU (6 GB)
- 512 Service and I/O nodes
- 200 Cabinets
- 710 TB total system memory
- Cray Gemini 3D Torus Interconnect
- 8.9 MW peak power – 8.3 avg.

X86 processor provides fast, single thread performance for control & communications

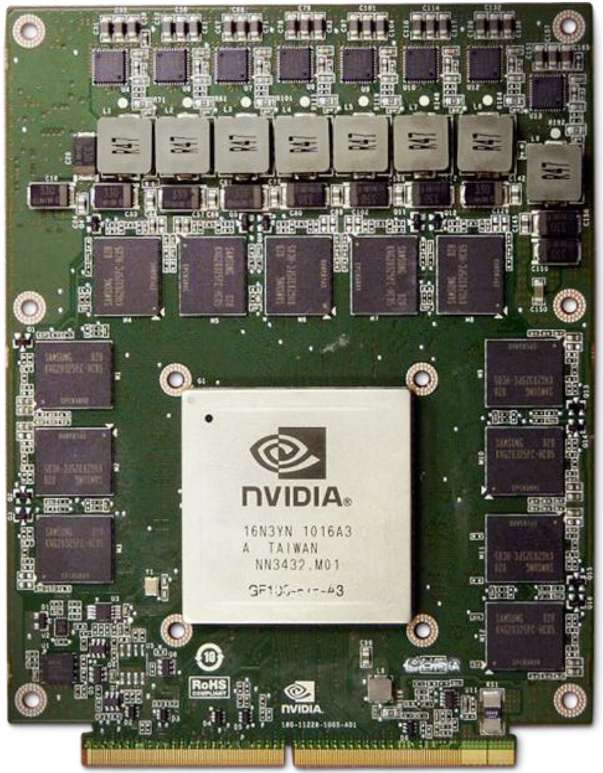


AMD Opteron 6274

- 16 cores
- 141 GFLOPs peak

GPUs are designed for extreme parallelism, performance & power efficiency

NVIDIA

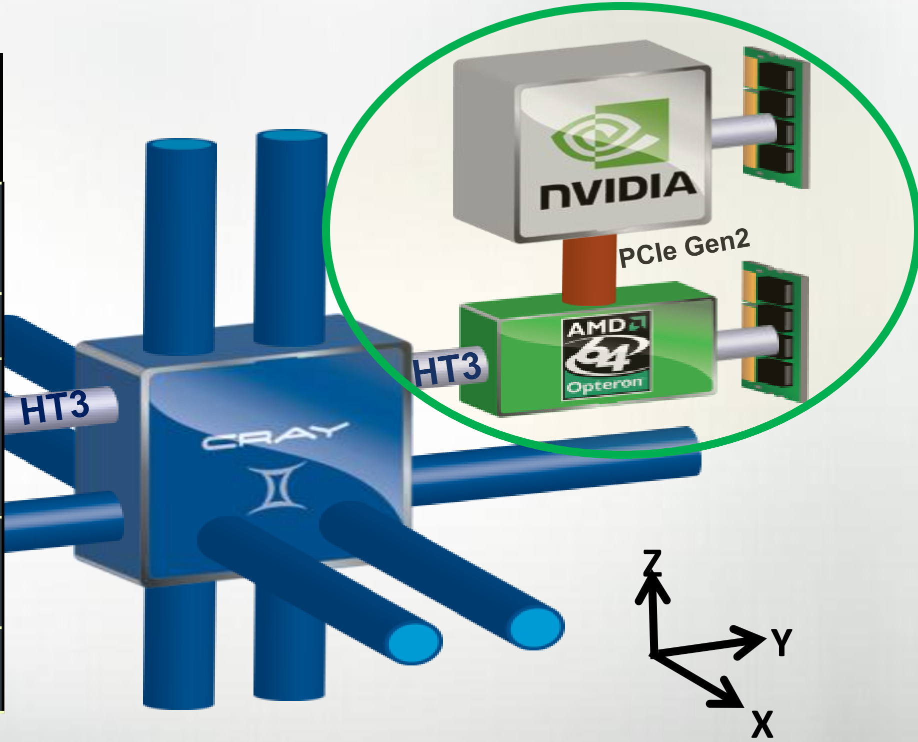


NVIDIA Tesla K20x

- 14 Streaming Multiprocessors
- 2,688 CUDA cores
- 1.31 TFLOPs peak (DP)
- 6 GB GDDR5 memory
- HPL: >2.0 GFLOPs per Watt
  (Titan full system measured power)

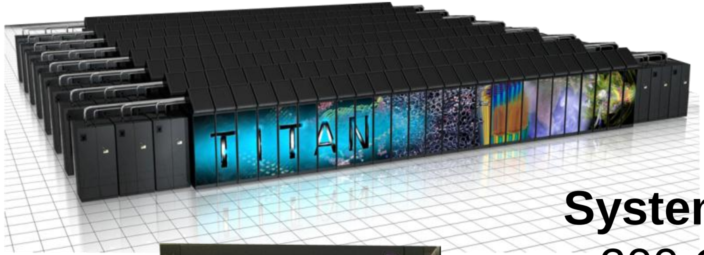
Cray XK7 Compute Node

XK7 Compute Node Characteristics
AMD Opteron 6274 16 core processor @ 141 GF
Tesla K20x @ 1311 GF
Host Memory 32GB 1600 MHz DDR3
Tesla K20x Memory 6GB GDDR5
Gemini High Speed Interconnect

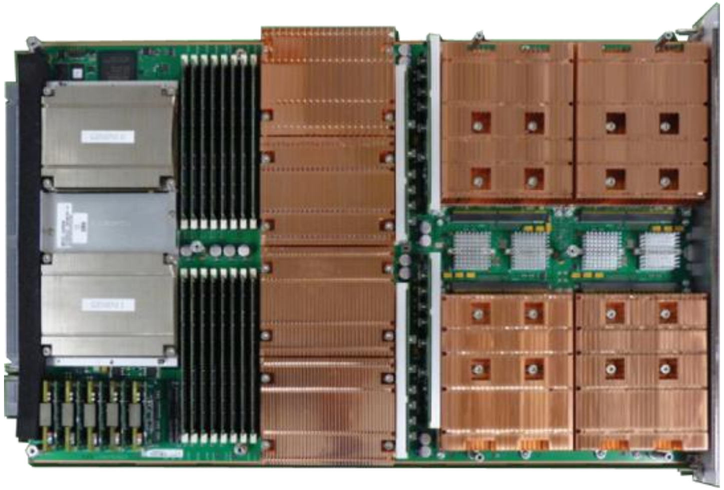


Slide courtesy of Cray, Inc.

Titan: Cray XK7 System

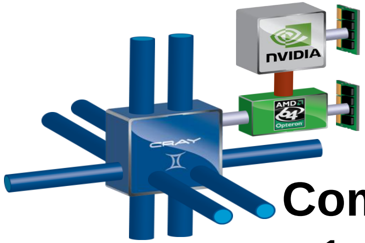


System:
200 Cabinets
18,688 Nodes
27 PF
710 TB



Cabinet:
24 Boards
96 Nodes
139 TF
3.6 TB

Board:
4 Compute Nodes
5.8 TF
152 GB



Compute Node:
1.45 TF
38 GB

Why GPUs? High Performance and Power Efficiency on a Path to Exascale

- Hierarchical parallelism – Improves scalability of applications
- Exposing more parallelism through code refactoring and source code directives
- Heterogeneous multi-core processor architecture – Use the right type of processor for each task.
- Data locality – Keep the data near the 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.

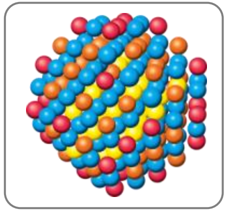
Hybrid Programming Model

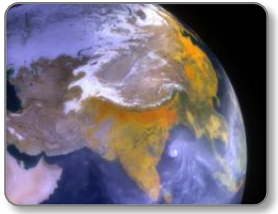
- On Jaguar, with 299,008 cores, we were seeing the limits of a single level of MPI scaling for most applications
- To take advantage of the vastly larger parallelism in Titan, users need to use hierarchical parallelism in their codes
  - Distributed memory: MPI, SHMEM, PGAS
  - Node Local: OpenMP, Pthreads,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

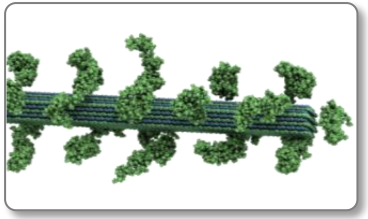
How do you program these nodes?

- **Compilers**
  - OpenACC is a set of compiler directives that allows the user to express hierarchical parallelism in the source code so that the compiler can generate parallel code for the target platform, be it GPU, MIC, or vector SIMD on CPU
  - Cray compiler supports XK7 nodes and is OpenACC compatible
  - CAPS HMPP compiler supports C, C++ and Fortran compilation for heterogeneous nodes with OpenACC support
  - PGI compiler supports OpenACC and CUDA Fortran
- **Tools**
  - Allinea DDT debugger scales to full system size and with ORNL support will be able to debug heterogeneous (x86/GPU) apps
  - ORNL has worked with the Vampir team at TUD to add support for profiling codes on heterogeneous nodes
  - CrayPAT and Cray Apprentice support XK6 programming

Early Science Applications on Titan





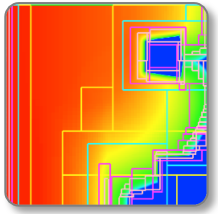


Material Science (WL-LSMS)
Role of material disorder, statistics, and fluctuations in nanoscale materials and systems.

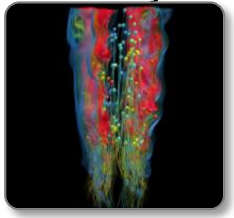
Climate Change (CAM-SE)
Answer questions about specific climate change adaptation and mitigation scenarios; realistically represent features like precipitation patterns/statistics and tropical storms.

Biofuels (LAMMPS)
A multiple capability molecular dynamics code.

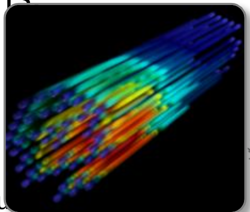
Astrophysics (NRDF)
Radiation transport – critical to astrophysics, laser fusion, combustion, atmospheric dynamics, and medical imaging.



Combustion (S3D)
Combustion simulations to enable the next generation of diesel/bio- fuels to burn more efficiently.



Nuclear Energy (Denovo)
Unprecedented high-fidelity radiation transport calculations that can be used in a variety of nuclear energy and technology applications.



Bu

How Effective are GPUs on Scalable Applications?

OLCF-3 Early Science Codes

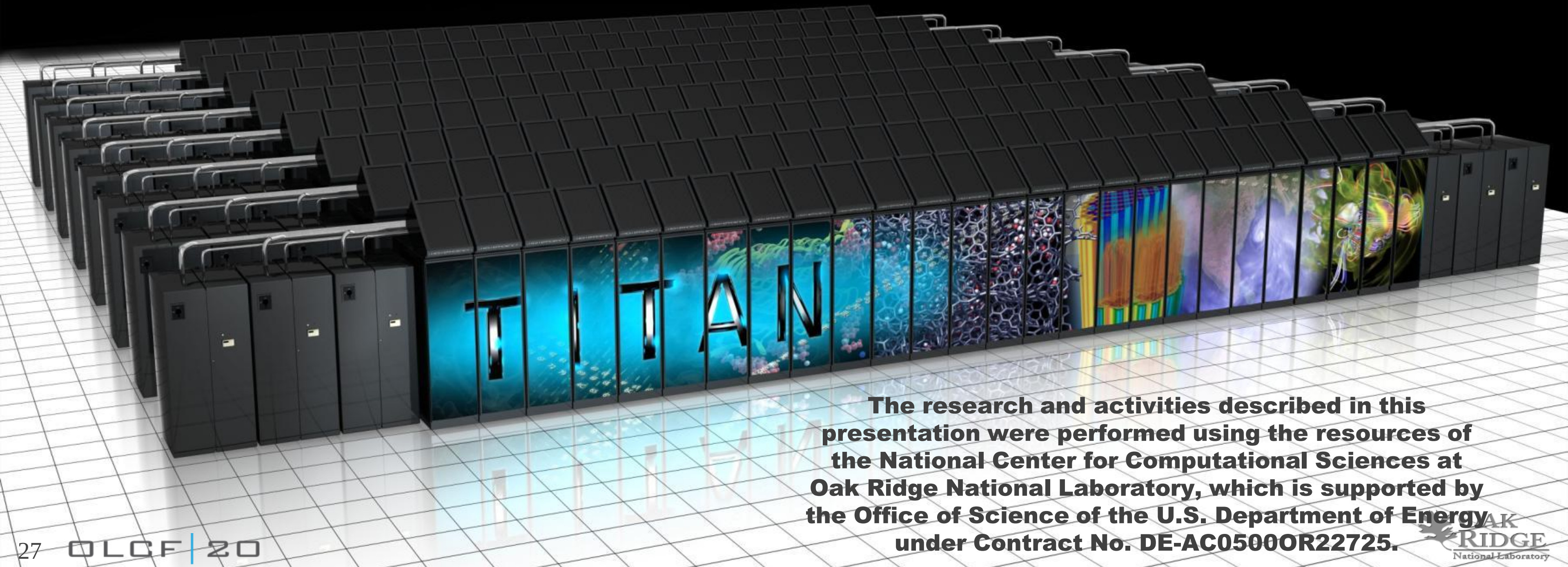
Very early performance measurements on Titan

	XK7 (w/ K20x) vs. XE6	Cray XK7: K20x GPU plus AMD 6274 CPU Cray XE6: Dual AMD 6274 and no GPU Cray XK6 w/o GPU: Single AMD 6274, no GPU
Application	**Performance Ratio**	**Comments**
S3D	1.8	• Turbulent combustion • **6% of Jaguar workload**
Denovo sweep	3.8	• Sweep kernel of 3D neutron transport for nuclear reactors • **2% of Jaguar workload**
LAMMPS	7.4* (mixed precision)	• High-performance molecular dynamics • **1% of Jaguar workload**
WL-LSMS	3.8	• Statistical mechanics of magnetic materials • **2% of Jaguar workload** • 2009 Gordon Bell Winner
CAM-SE	1.8* (estimate)	• Community atmosphere model • **1% of Jaguar workload**

Questions?

BlandAS@ornl.gov

Want to join our team? ORNL is hiring. Contact us at http://jobs.ornl.gov



The research and activities described in this presentation were performed using the resources of the National Center for Computational Sciences at Oak Ridge National Laboratory, which is supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC0500OR22725.