NCSA

All New & Improved ^

Blue Waters

Overview of the Sustained Petaflop Blue Waters System

Thom Dunning, Bill Kramer, Marc Snir, Bill Gropp and Wen-mei Hwu

National Center for Supercomputing Applications
University of Illinois at Urbana-Champaign

National Center for Supercomputing Applications

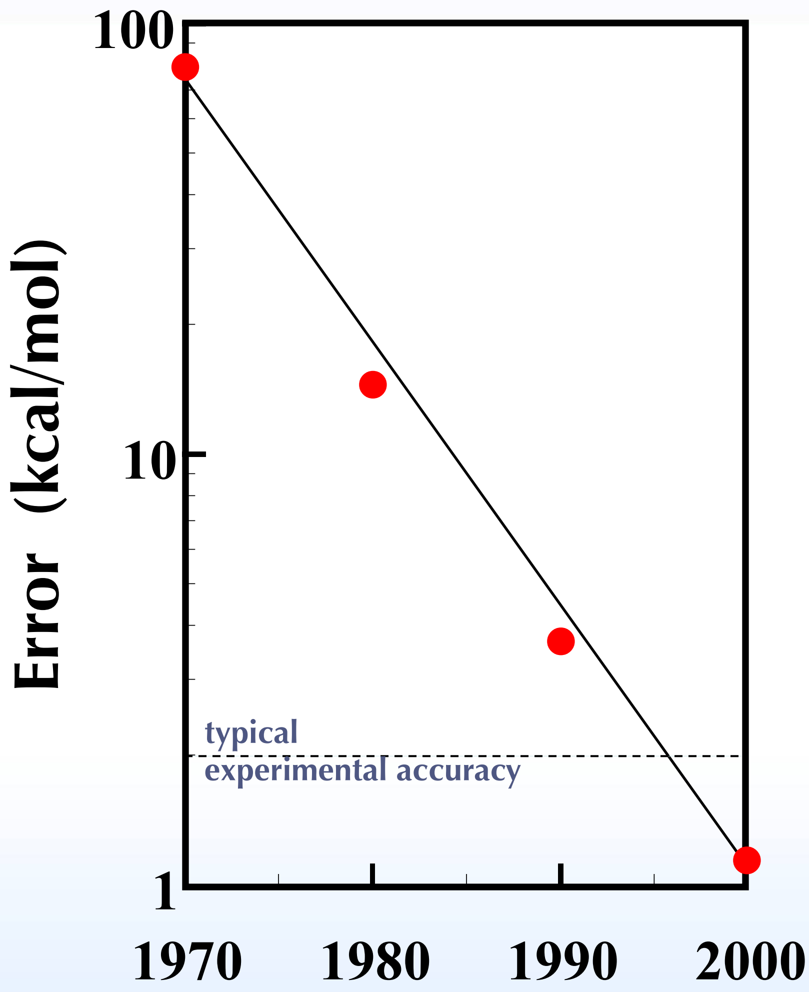
Established in 1986 as a partnership between the National Science Foundation, the University of Illinois at Urbana-Champaign, and the State of Illinois


- **NCSA**
  - R&D unit of the University of Illinois at Urbana-Champaign
  - One of original five NSF-funded supercomputing centers
    - **Mission**: Provide state-of-the-art computing capabilities (both hardware and software) to nation's scientists and engineers
- **The Numbers**
  - Approximately 250 staff (200 technical/professional staff)
  - Two major facilities (NCSA Building, NPCF)
  - Deploying NSF's most powerful computing system: Blue Waters
  - Managing NSF's national cyberinfrastructure: XSEDE

Need for Increased Computing Power

Increased computing power enables scientists and engineers to extend the accuracy and scope of their computational simulations (with examples from molecular science)

Increasing Accuracy of Simulations



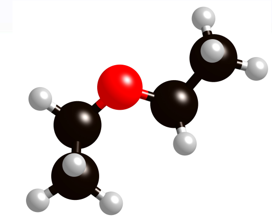
- Bond energies critical for describing many chemical phenomena, e.g., combustion
- Difficult to measure in the laboratory
- Accuracy of calculated bond energies increased dramatically from 1970-2000
- Due to advances in
  - Theoretical methodology
  - Computational techniques
  - Computing technology

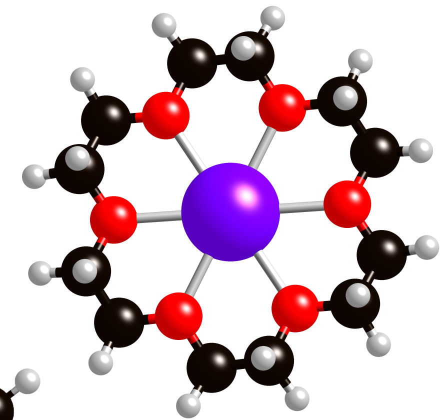
Increasing the Scope of Simulations

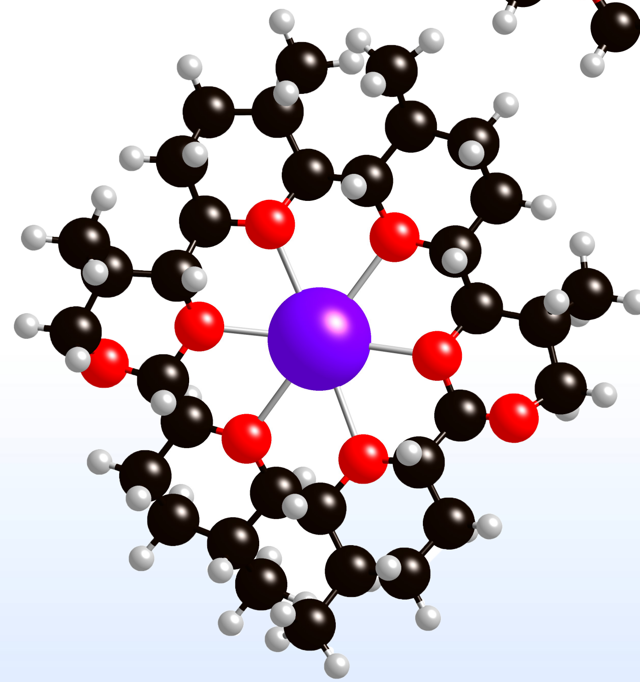
- **In 1990**
  - Model separations agents, *e.g.*, ethyl ether

– **In 2000**
  - Prototype separation agents, *e.g.*, 18-crown-6

- **In 2005**
  - Real-world separations agents, *e.g.*, Still's crown ether







Figures courtesy of B. Hay, Oak Ridge National Laboratory

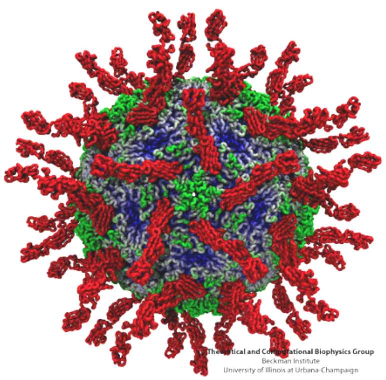
Petascale Applications

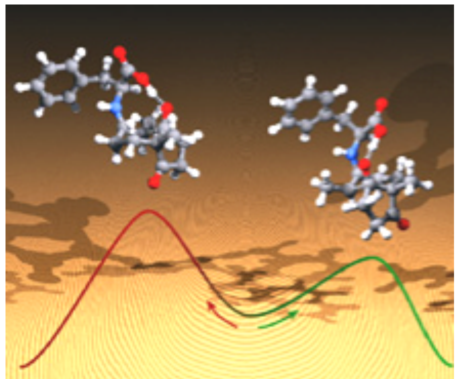
There are a broad range of applications that need petascale computing capabilities—from astrophysics and biophysics to climate and earthquake modeling; materials, molecular and nanoscale science; and weather prediction

Advancing Science and Engineering

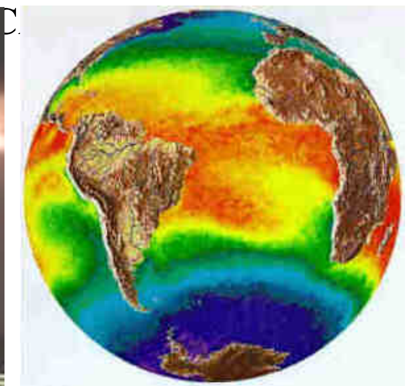
Blue Waters will enable breakthroughs in a broad range of science and engineering disciplines:

MOLECULAR SCIENCE



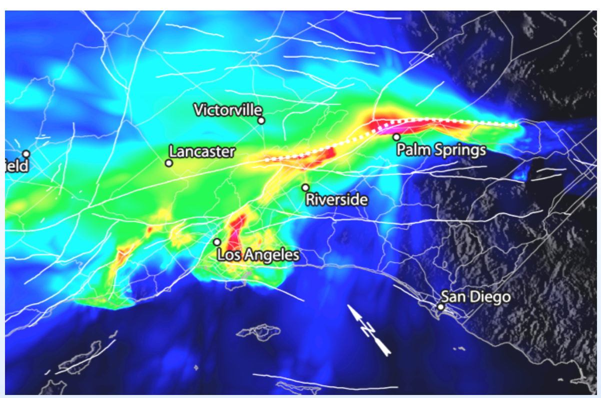


WEATHER & CLIMATE

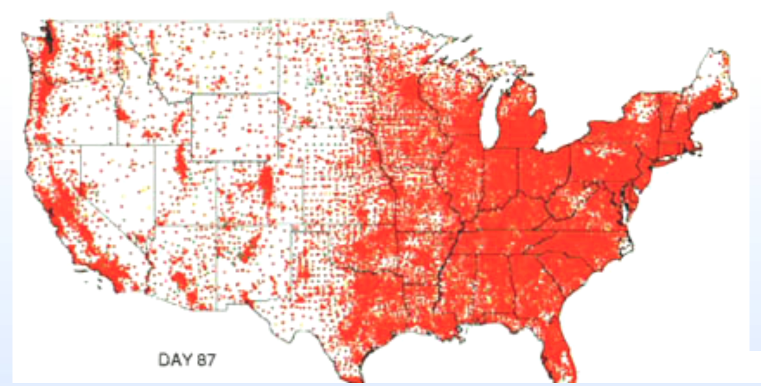


ASTRONOMY

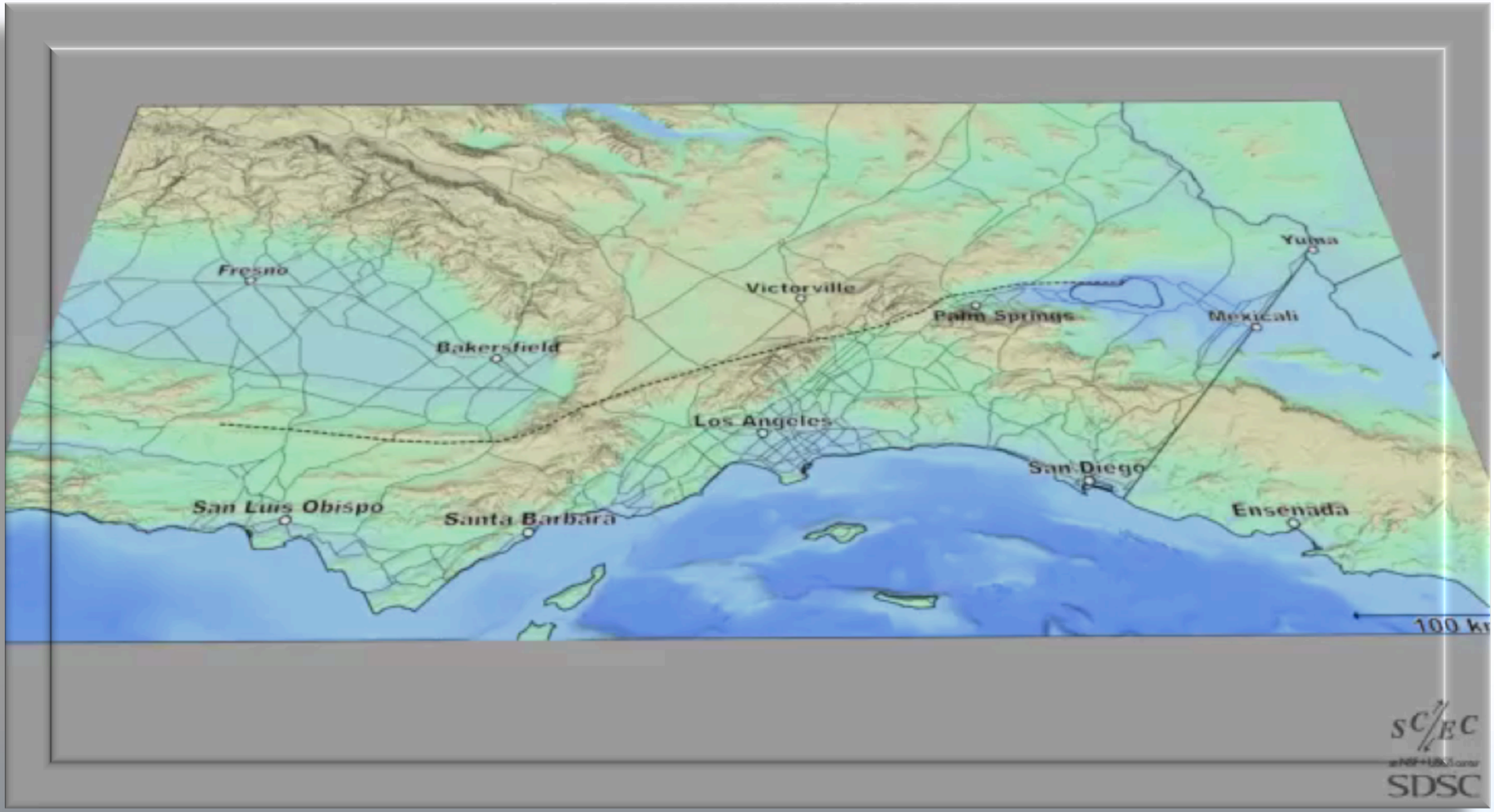
EARTH



HEALTH

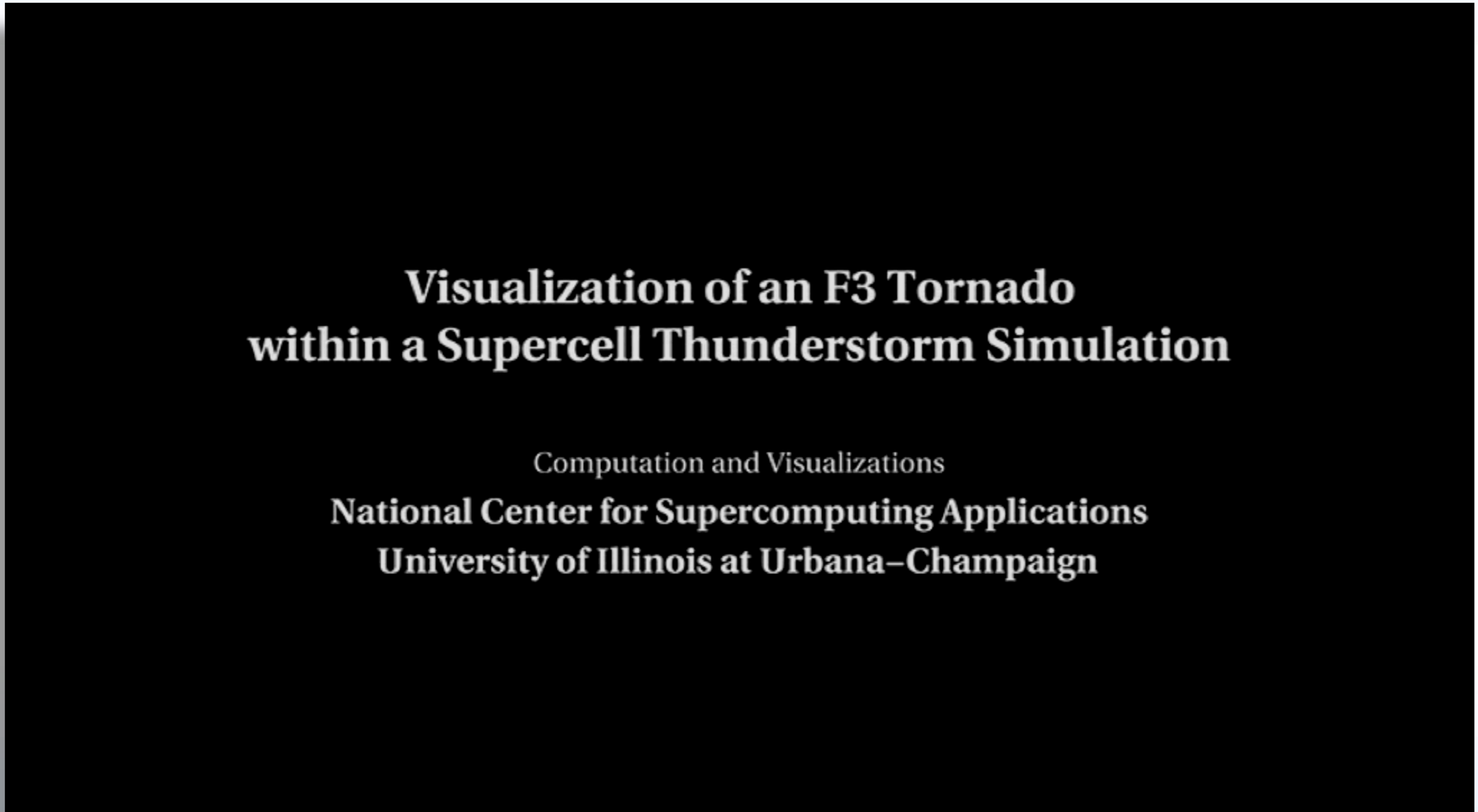


Modeling Earthquakes on San Andreas Fault

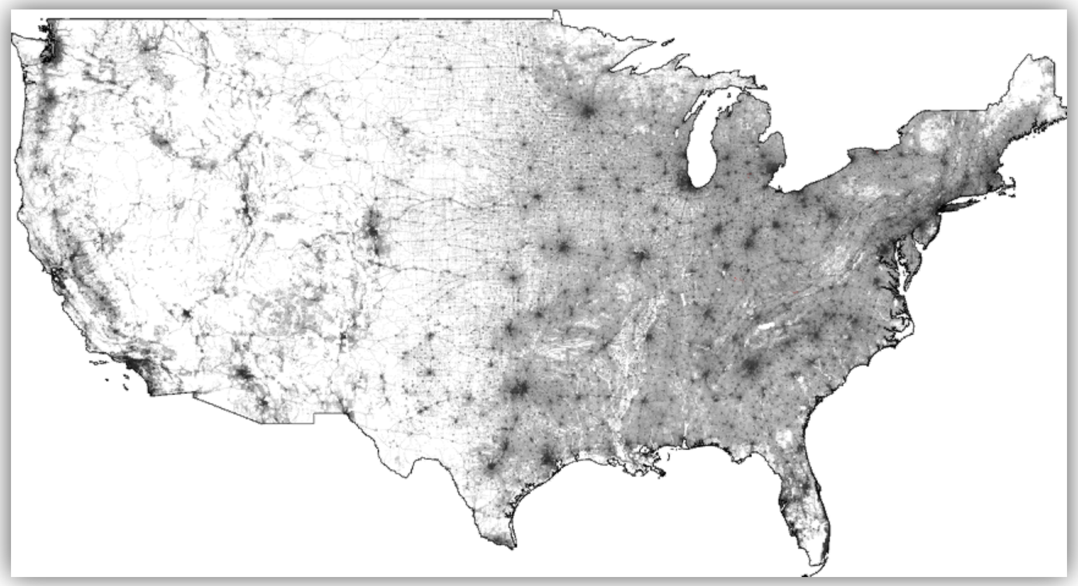


Video courtesy of T. Jordan, Southern California Earthquake Center

Simulating the Birth of Tornados



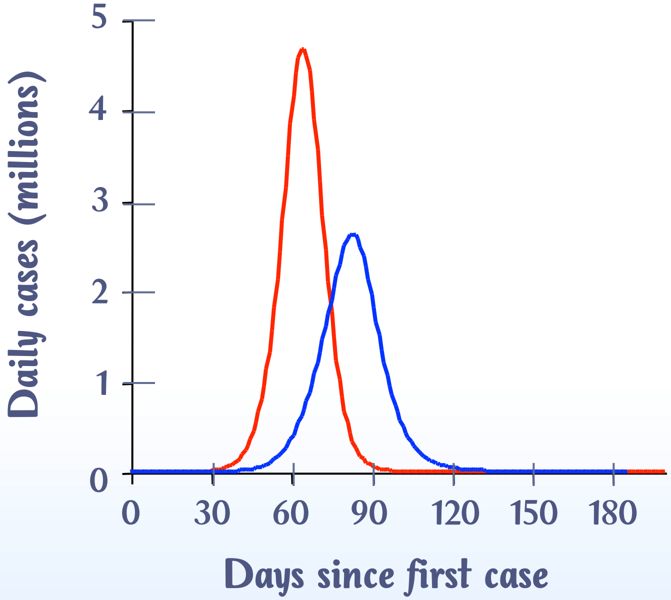
Modeling Spread of Infectious Diseases



Video courtsey of N. Ferguson,
Imperial College, London

No Intervention

Intervention: next-day treatment of 90% of cases with anti-virals, school closures, 50% household quarantine.



Blue Waters

Blue Waters is designed to handle the most compute-, memory- and data-intensive simulations and analyses in science and engineering

Input from the Research Community

- **Maximum Core Performance**

  … to minimize number of cores needed for a given performance level, lessen impact of sections of code with limited scalability

- **Low Latency, High Bandwidth Interconnect**

  … to enable science and engineering applications to scale to tens to hundreds of thousands of cores

- **Large, Fast Memories**

  … to solve the most memory-intensive problems

- **Large, Fast I/O System and Data Archive**

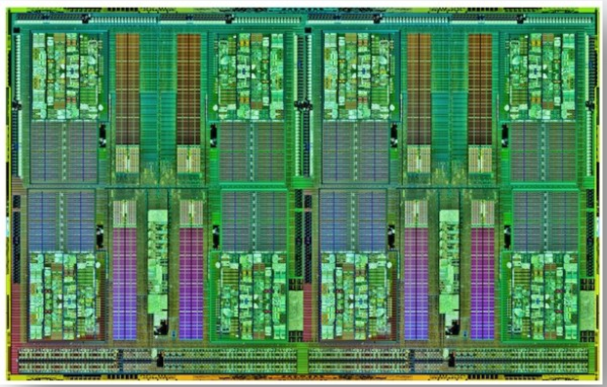
  … to solve the most data-intensive problems

- **Innovative Computing Technologies**

  … to explore the use of innovative computing technologies in science and engineering applications

New in 2011!

Heart of Blue Waters: Two New Chips

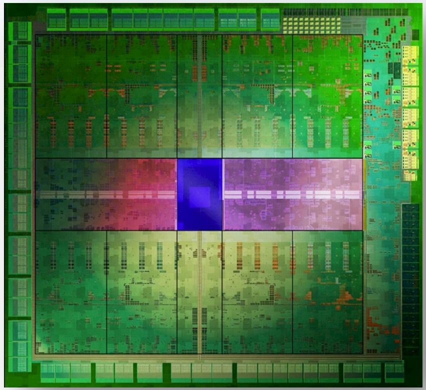


AMD Interlagos

156.8 GF peak performance

Features:

- 2.3-2.6 GHz
- 8 core modules, 16 threads
- On-chip Caches
  - L1 (I:8x64KB; D:16x16KB)
  - L2 (8x2MB)
- Memory Subsystem
  - Four memory channels
  - 51.2 GB/s bandwidth



NVIDIA Kepler

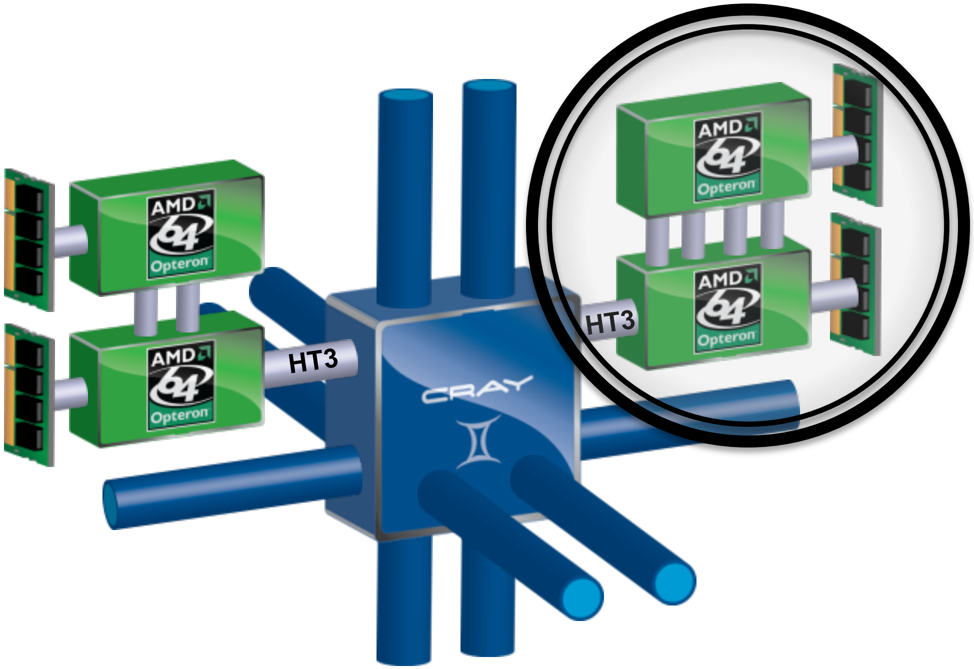
>1,000 GF peak performance

Features:

- 15 Streaming multiprocessors (SMX)
  - SMX: 192 sp CUDA cores, 64 dp units, 32 special function units
  - L1 caches/shared mem (64KB, 48KB)
  - L2 cache (1536KB)
- Memory subsystem
  - Six memory channels
  - 180 GB/s bandwidth

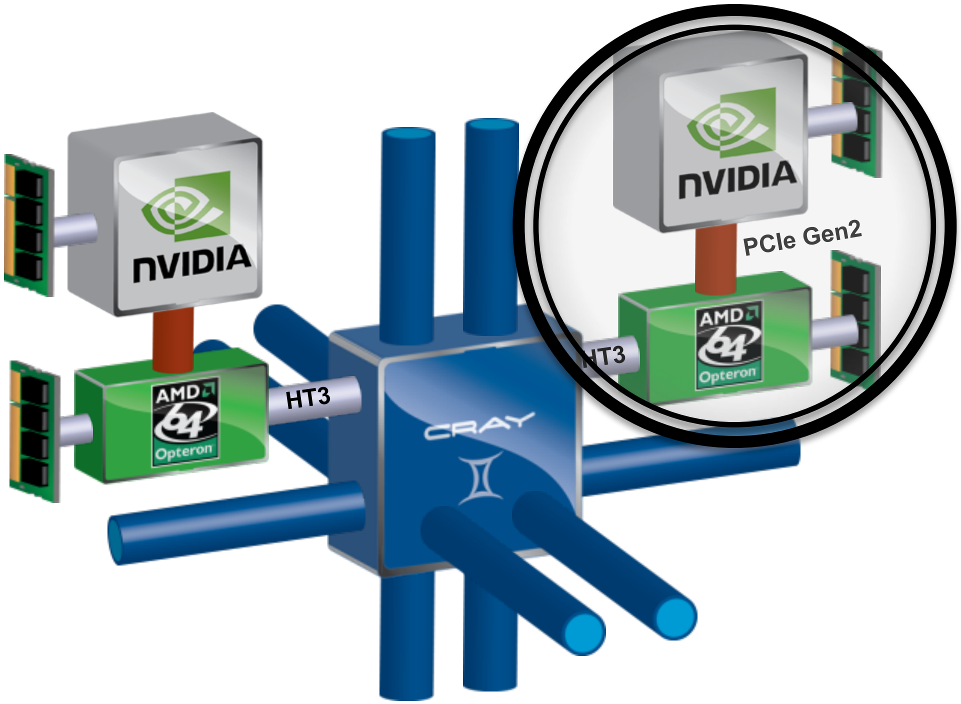
Cray XE6 Nodes

- **Dual-socket Node**
  - Two AMD Interlagos chips
    - 16 core modules, 64 threads
    - 313.6 GFs peak performance
    - 64 GBs memory
      - 102 GB/sec memory bandwidth
  - Gemini Interconnect
    - Router chip & network interface
    - Injection Bandwidth (peak)
      - 9.6 GB/sec per direction



Blue Waters contains 22,640 Cray XE6 compute nodes.

Cray XK7 Nodes

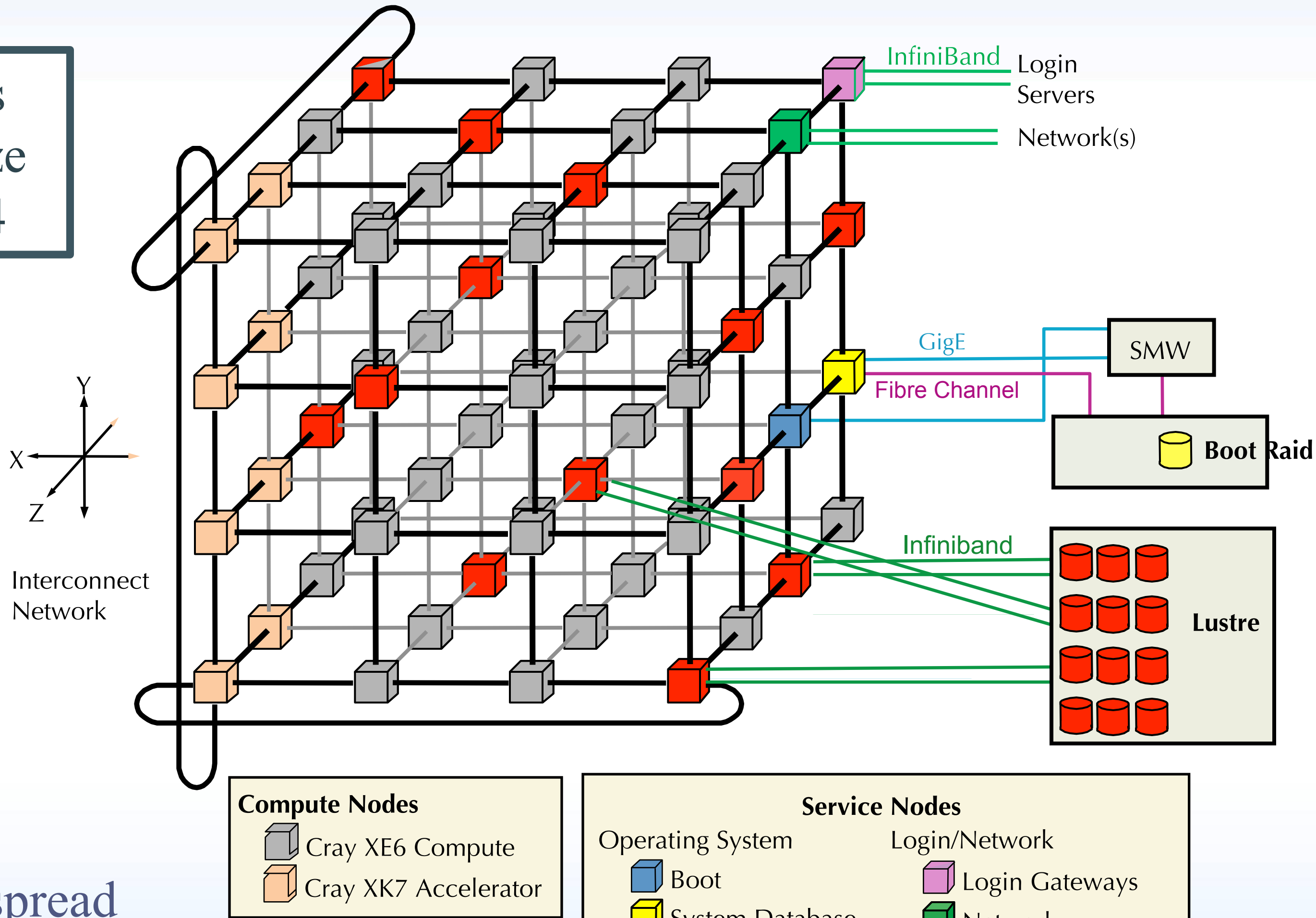


- **Dual-socket Node**
  - One AMD Interlagos chip
    - *Same as XE6 nodes*
  - One NVIDIA Kepler chip
    - >1 TF peak performance
    - 6 GBs GDDR5 memory
      - 180 GB/sec bandwidth
  - Gemini Interconnect
    - *Same as XE6 nodes*

Blue Waters contains 3,072 Cray XK7 compute nodes.

Gemini Interconnect Network

Blue Waters
3D Torus Size
23 x 24 x 24



Service Nodes spread throughout the torus

Blue Waters Disk Subsystem

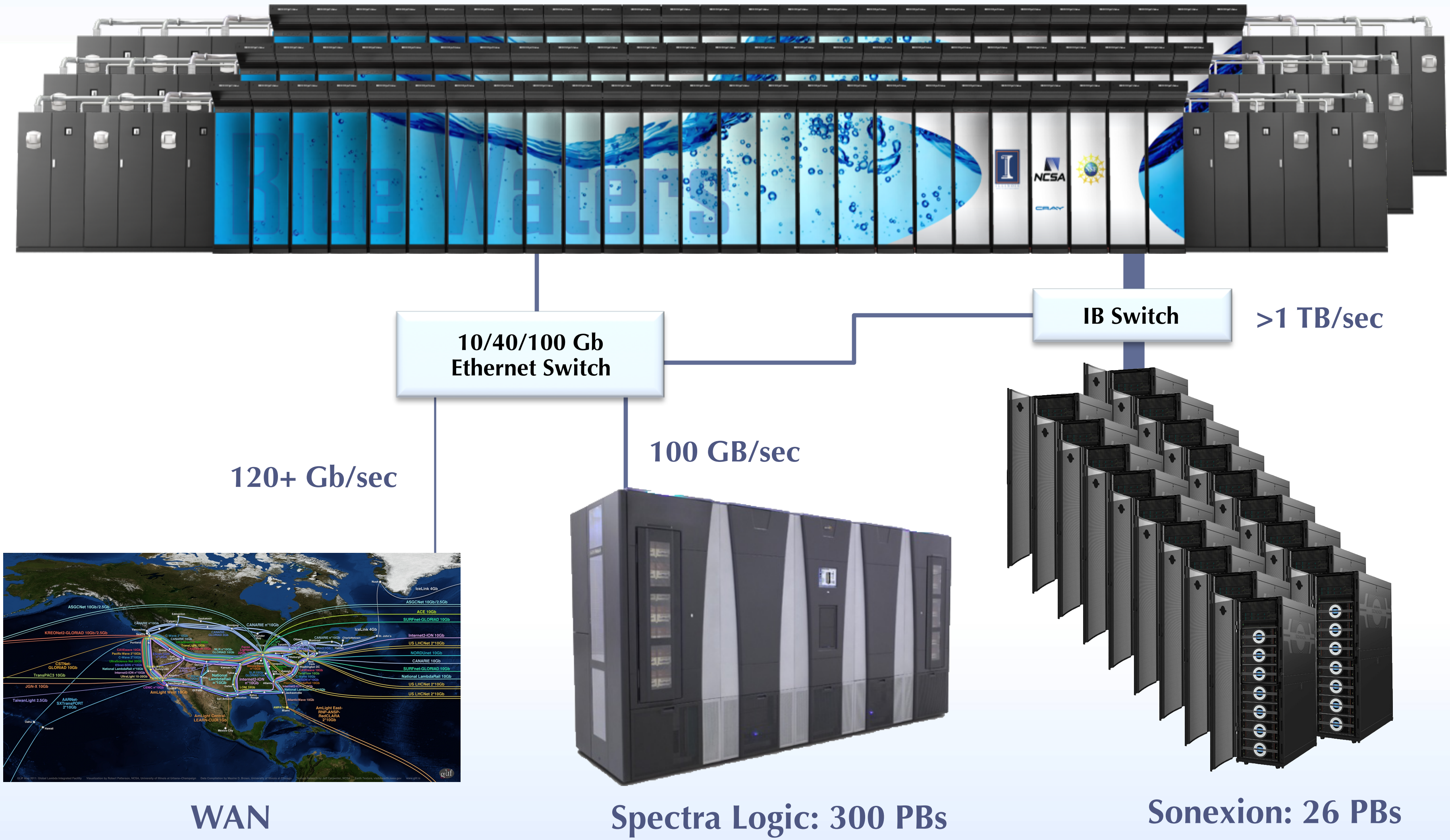
CRAY
THE SUPERCOMPUTER COMPANY

- **Cray Sonexion 1600**
  - Lustre file system
  - Reliable, Modular, Scalable
  - Fully integrated
    - Servers
    - Disk drives (Scalable Storage Units)
    - QDR Infiniband switches
  - Hierarchical monitoring
- **Blue Waters Disk Subsystem**
  - Capacity: 34.6 PBs (*raw*), 25.9 PBs (*usable*)
  - Bandwidth: >1 TB/s (*sustained*)

Blue Waters Archive System

- **Spectra Logic T-Finity**
  - Dual-arm robotic tape libraries
  - High availability and reliability, with built-in redundancy
- **Blue Waters Archive**
  - Capacity: 380 PBs (*raw*), 300 PBs (*usable*)
  - Bandwidth: 100 GB/sec (*sustained*)
  - RAIT for increased reliability

Blue Waters Computing System



Blue Waters Software Environment

Languages
- Fortran
- C
- C++
- Python
- UPC

Compilers
- Cray Compiling Environment (CCE)
- PGI
- GNU

Programming Models
- Distributed Memory (Cray MPT)
  - MPI
  - SHMEM
- Shared Memory
  - OpenMP 3.0
- PGAS & Global View
  - UPC (CCE)
  - CAF (CCE)

IO Libraries
- NetCDF
- HDF5
- ADIOS

Tools
- Environment setup
  - Modules
- Debugging Support Tools
  - Fast Track Debugger (CCE w/ DDT)
  - Abnormal Termination Processing
  - STAT
  - Cray Comparative Debugger#
- Data Transfer
  - globus online
  - HPSS

Scientific Libraries
- LAPACK
- ScaLAPACK
- BLAS (libgoto)
- Iterative Refinement Toolkit
- Cray Adaptive FFTs (CRAFFT)
- FFTW
- Cray PETSc (with CASK)
- Cray Trilinos (with CASK)

Performance Analysis
- Cray Performance Monitoring and Analysis Tool
- PAPI
- PerfSuite
- Tau

Debuggers
- Allinea DDT
- lgdb

Prog. Env.
- Eclipse
- Traditional

Resource Manager
- Adaptive Computing

Visualization
- VisIt
- Paraview
- YT

Cray Linux Environment (CLE)/SUSE Linux

Legend:

Cray developed	Licensed ISV Software	NCSA supported
Under development	3rd Party Package	Cray added value to 3rd party

Blue Waters & 京 Computing Systems

System Attribute	NCSA Blue Waters	RIKEN 京 (#2)
Vendor(s)	Cray/AMD/NVIDIA	Fujitsu
Processors	Interlagos/Kepler	SPARC64 VIIIfx
Total Peak Performance (PF)	**11.6**	**11.3**
Total Peak Performance (CPU/GPU)	**7.6/4.0**	**11.3/0.0**
Number of CPU Chips (8 cores/chip)	48,352	88,128
Number of GPU Chips	3,072	0
Amount of CPU Memory (TB)	**1,518**	**1,410**
Interconnect	3-D Torus	6-D Torus
Amount of On-line Disk Storage (PB)	**26**	**11/30**
Sustained Disk Transfer (TB/sec)	**>1**	?
Amount of Archival Storage	**300**	?
Sustained Tape Transfer (GB/sec)	**100**	?

Blue Waters & Titan Computing Systems

System Attribute	NCSA Blue Waters	ORNL Titan
Vendor(s)	Cray/AMD/NVIDIA	Cray/AMD/NVIDIA
Processors	Interlagos/Kepler	Interlagos/Kepler
Total Peak Performance (PF)	**11.6**	**~20**
Total Peak Performance (CPU/GPU)	**7.6/4.0**	**3/17**
Number of CPU Chips (8 cores/chip)	48,352	18,688
Number of GPU Chips	3,072	14,592
Amount of CPU Memory (TB)	**1,518**	**688**
Interconnect	3-D Torus	3-D Torus
Amount of On-line Disk Storage (PB)	**26**	**20(?)**
Sustained Disk Transfer (TB/sec)	**>1**	**0.4–0.7**
Amount of Archival Storage	**300**	**15-30**
Sustained Tape Transfer (GB/sec)	**100**	**7**

Petascale Computing Facility

Partners

EYP MCF/
Gensler
IBM
Yahoo!

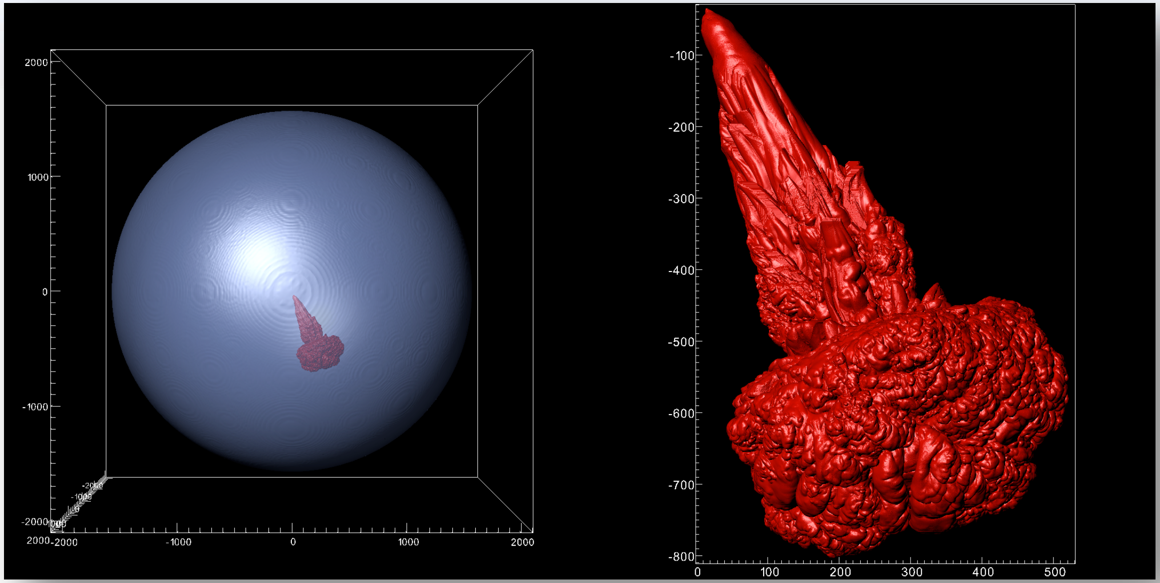
- **Modern Data Center**
  - 90,000+ ft^2 total
  - 30,000 ft^2 raised floor
    20,000 ft^2 machine room gallery
- **Energy Efficiency**
  - LEED certified Gold
  - Power Utilization Efficiency = 1.1–1.2

Blue Waters Early Science System

- **BW-ESS Configuration**
  - 1.4+ PFs (peak)
    - 48 cabinets (4,512 XE6 compute nodes)
    - 96 service nodes
  - 2 PBs Sonexion Lustre storage appliance

- **BW-ESS Projects**
  - **Biomolecular Physics**—K. Schulten, University of Illinois at Urbana-Champaign
  - **Cosmology**—B. O'Shea, Michigan State University
  - **Climate Change**—D. Wuebbles, University of Illinois at Urbana-Champaign
  - **Lattice QCD**—R. Sugar, University of California, Santa Barbara
  - **Plasma Physics**—H. Karimabadi, University of California, San Diego
  - **Supernovae**—S. Woosley, University of California Observatories
  - **Severe Weather—**R. Wilhelmson, University of Illinois
  - **High Resolution/Fidelity Climate—**C. Stan, Center for Ocean-Land-Atmospheric Studies (COLA)
  - **Complex Turbulence**—P. K. Yeung, Georgia Tech
  - **Turbulent Stellar Hydrodynamics—**P. Woodward, University of Minnesota

Modeling Type 1a Supernovae



- **PI: S. Woosley, University of California Observatories**
- **Details**
  - Off-center ignition of Type 1a Supernovae, 1 second duration
  - Codes: MAESTRO and CASTRO
  - Used 68 million core hours
  - Produced 45 TBs of data

Current Build Status of Blue Waters

- **All compute cabinets, XE6/XK7, installed and being tested**
- **All disks (36 racks) delivered and upgraded to 1600 controllers and being tested**
- **Near-line storage: one Spectra Logic library (with 4000 slots for 4TB media) installed and being used for HPSS and Cray software testing**

Sustained Petascale Performance

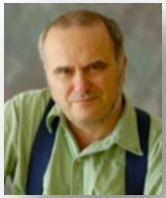
- **Linpack & Top500**
  - Limited means of determining true performance of computers
- **Sustained Petascale Performance (SPP) Metric**
  - Similar to NERSC Sustained System Performance metric
  - Time-to-solution metric for end-to-end problem (pre/post processing, processing, I/O)
  - Coverage of representative science areas, algorithmic methods and core counts
- **Details**
  - Original NSF Benchmarks
    - Full Size: QCD (MILC), Turbulence (PNSDNS), Molecular Dynamics (NAMD)
    - Modest Size: MILC, Paratec, WRF
  - Science Teams Applications
    - **NAMD**: molecular dynamics; **MILC**, **Chroma**: Lattice Quantum Chromodynamics; **VPIC**, **SPECFEM3D**: Geophysical Science; **WRF**: Atmospheric Science; **PPM**: Astrophysics; **NWCHEM**, **GAMESS**: Computational Chemistry; **QMCPACK**: Materials Science
    - At least three Science Team benchmarks must be run at full scale

Illinois' Blue Waters Team

PI & Co-PIs


Dunning


Kramer

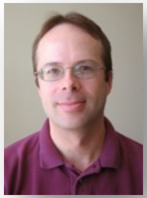

Snir


Gropp

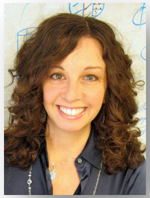

Hwu

Task Leads


Beldica
Proj. Mgmt.


Bode
Software


Butler
Storage


Glotzer
Virtual School


Cappello
Reliability


Kale
Apps Simulations


Lathrop
Education


Durbin
Facilities


Giles
Industry


Olson
Proj. Mgmt.


Panoff
UG Education


Semeraro
Visualization


Enos
Ops Transition

Thank You!