NCSA

Blue Waters

An Extraordinary Computer to Enable Extraordinary Research

Thom Dunning

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

National Center for Supercomputing Applications

A brief description of NCSA, which celebrates its Silver Anniversary in 2011

National Center for Supercomputing Applications

- **Founded in 1986**
  - 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)
  - Three major computing systems (Abe, Lincoln, Ember)

NCSA's R&D Programs

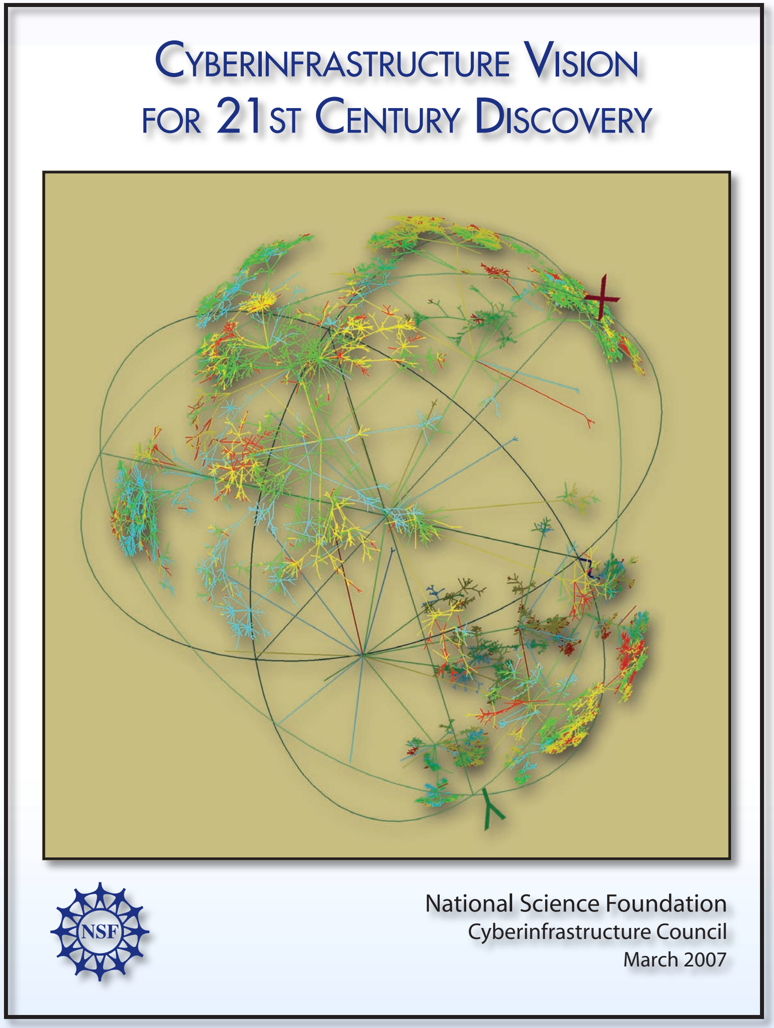
- **Director's Office**
  - **Major Programs**
    - Extreme-scale Computing
    - Grid Computing
    - Data-driven Computing
  - **Laboratories**
    - Advanced Visualization Laboratory
    - Innovative Systems Laboratory
    - Cybersecurity Laboratory
    - Laboratory for Advancing STEM Education

Blue Waters

An extraordinary computer
to enable extraordinary research

NSF's Strategy for High-end Computing

- **Three Resource Levels**
  - **Track 3**: University owned and operated
  - **Track 2**: Several NSF-funded supercomputer & specialized computing centers (TeraGrid)
  - **Track 1**: NSF-funded leading-edge computer center
- **Computing Resources**
  - **Track 3**: 10s–100s TF
  - **Track 2**: 500–1,000 TF
  - **Track 1**: *see following slide*



NSF's Track 1 Solicitation

"The petascale HPC environment will enable investigations of computationally challenging problems that require computing systems capable of delivering **sustained performance approaching 10^{15} floating point operations per second** (petaflops) on real applications, that consume **large amounts of memory**, and/or that work with **very large data sets**."

Leadership-Class System Acquisition - Creating a Petascale Computing Environment for Science and Engineering

NSF 06-573

Blue Waters Project

Input from Scientific Community

- **D. Baker, University of Washington**
  Protein structure refinement and determination
- **M. Campanelli, RIT**
  Computational relativity and gravitation
- **D. Ceperley, UIUC**
  Quantum Monte Carlo molecular dynamics
- **J. P. Draayer, LSU**
  Ab initio nuclear structure calculations
- **P. Fussell, Boeing**
  Aircraft design optimization
- **C. C. Goodrich**
  Space weather modeling
- **M. Gordon, T. Windus, Iowa State University**
  Electronic structure of molecules
- **S. Gottlieb, Indiana University**
  Lattice quantum chromodynamics
- **V. Govindaraju**
  Image processing and feature extraction
- **M. L. Klein, University of Pennsylvania**
  Biophysical and materials simulations
- **J. B. Klemp *et al.*, NCAR**
  Weather forecasting/hurricane modeling
- **R. Luettich, University of North Carolina**
  Coastal circulation and storm surge modeling
- **W. K. Liu, Northwestern University**
  Multiscale materials simulations
- **M. Maxey, Brown University**
  Multiphase turbulent flow in channels
- **S. McKee, University of Michigan**
  Analysis of ATLAS data
- **M. L. Norman, UCSD**
  Simulations in astrophysics and cosmology
- **J. P. Ostriker, Princeton University**
  Virtual universe
- **J. P. Schaefer, LSST Corporation**
  Analysis of LSST datasets
- **P. Spentzouris, Fermilab**
  Design of new accelerators
- **W. M. Tang, Princeton University**
  Simulation of fine-scale plasma turbulence
- **A. W. Thomas, D. Richards, Jefferson Lab**
  Lattice QCD for hadronic and nuclear physics
- **J. Tromp, Caltech/Princeton**
  Global and regional seismic wave propagation
- **P. R. Woodward, University of Minnesota**
  Astrophysical fluid dynamics

Requested Attributes of Petascale System

- **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

- **Reliable Operation**

  …to enable the solution of Grand Challenge problems

November 24, 20006

IBM wins DARPA funding

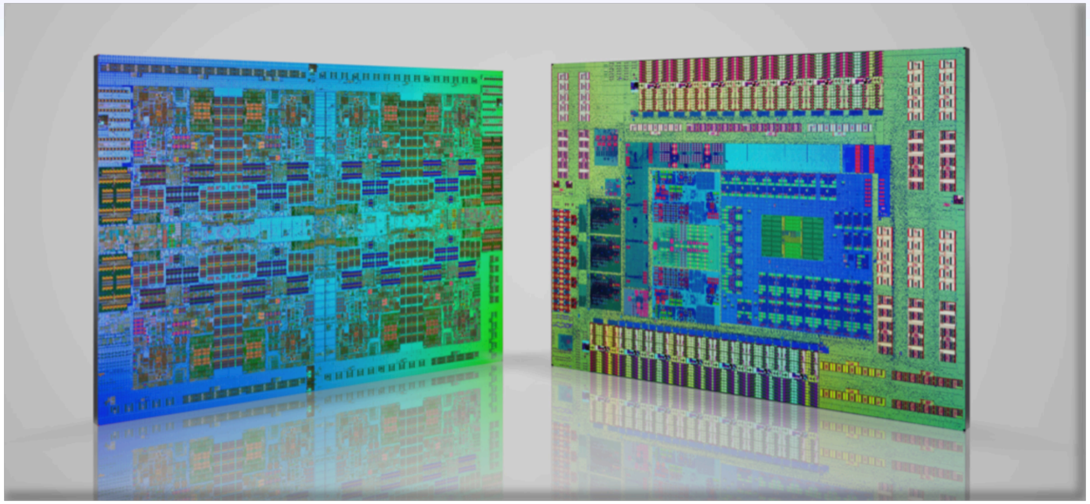
HPCS program allows IBM to pursue a vision of petascale computing systems

IBM's Programmable Easy-to-use Reliable Computing System (PERCS) was selected by the Defense Advanced Research Projects Agency (DARPA) as one of two system designs to be developed and demonstrated as part of phase III of the High Productivity Computing Systems program (HPCS). ***Such designs must support the eventual scaling of sustained computation to 10 petaflops and a software environment that enables domain experts to effectively use that computing power***. These capabilities are required by HPCS to meet the need for commercially successful petascale computing systems for high-end users in government, science and industry in 2010. …

PERCS will meet these goals with a scalable system based on future POWER series technologies. The PERCS program will substantially increase the research and development activities in IBM technologies planned for 2010 and beyond. These will enable IBM to meet the HPCS goals and enhance the capabilities of IBM's line of business systems. This will entail IBM making significant investments in the next generation of [a number of] technologies.

…

Heart of Blue Waters: Two New Chips



IBM Power7 Chip

Up to 256 GF peak performance

3.5–4.0+ GHz
Up to 8 cores, 32 threads
Caches
- L1 (2x64 KB), L2 (256 KB), L3 (32 MB)

Memory Subsystem
- Two memory controllers
- 128 GB/s memory bandwidth

IBM PERCS Hub Chip

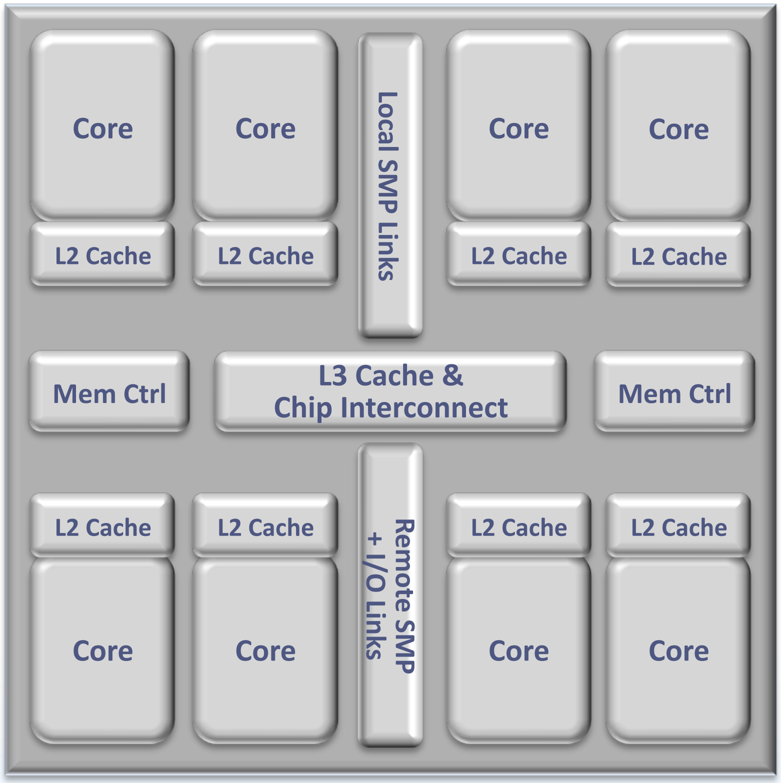
1.128 TB/s total bandwidth

Connections:

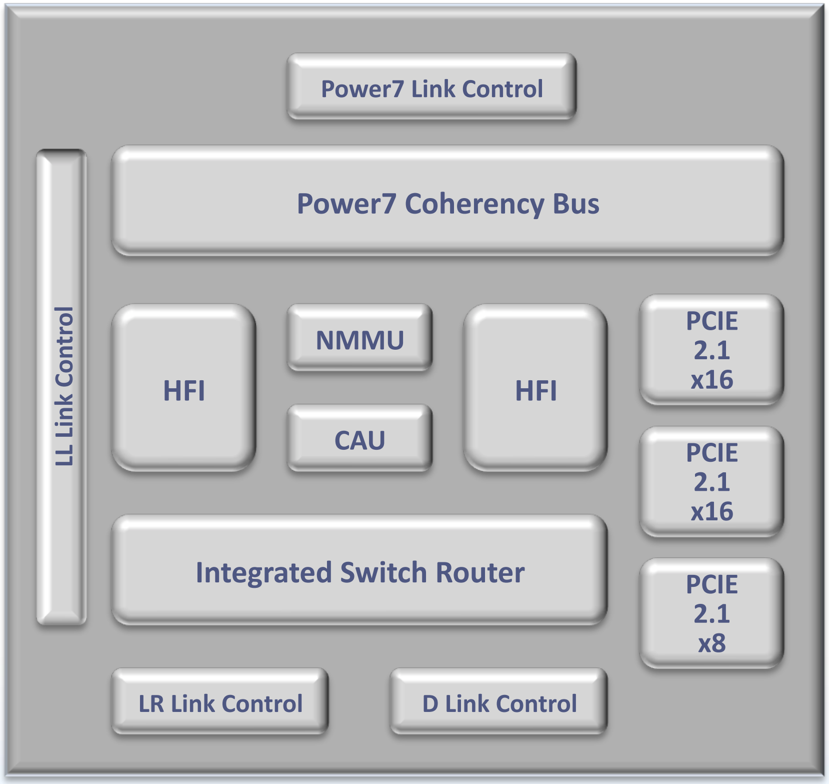
192	GB/s QCM connection
336	GB/s to QCMs in same drawer
240	GB/s to QCMs to other drawers in supernode
320	GB/s to other supernodes
40	GB/s general purpose I/O

Block Diagrams: Power7 and Hub Chips

Power7



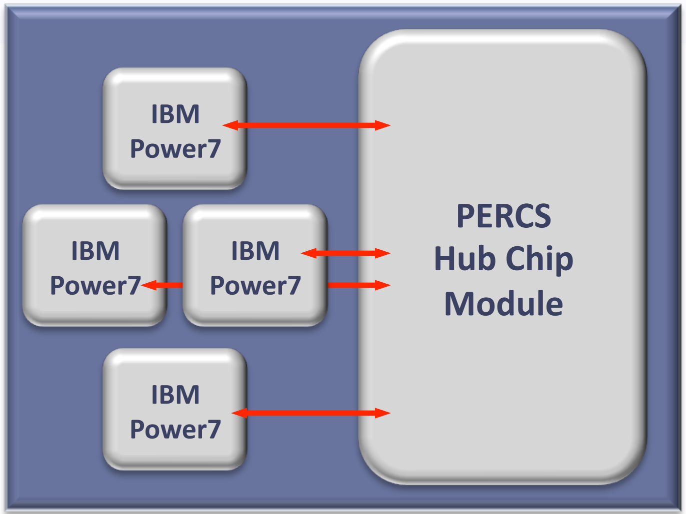
PERCS Hub Chip

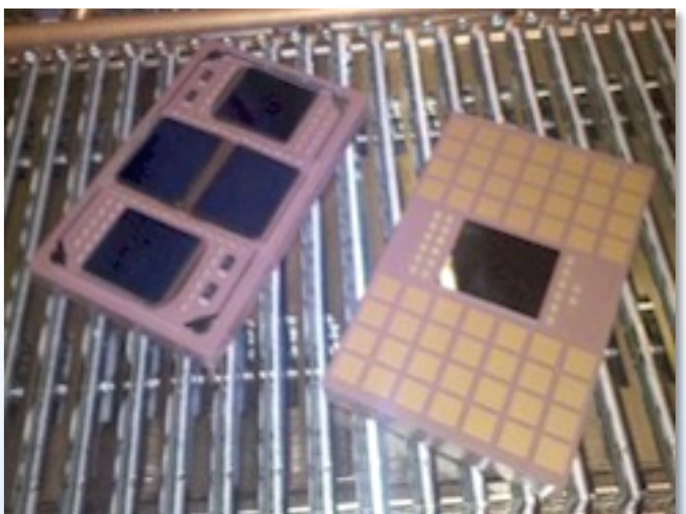


"Power7: IBM's Next Generation Server Processor," R. Kalla, B. Sinharoy, W. J. Starke, and M. Floyd, *IEEE Micro* **30**, 7 (2010)

"The PERSC High-Performance Interconnect," B. Arimilli *et al.*, to be published

Quad-chip Module (QCM)

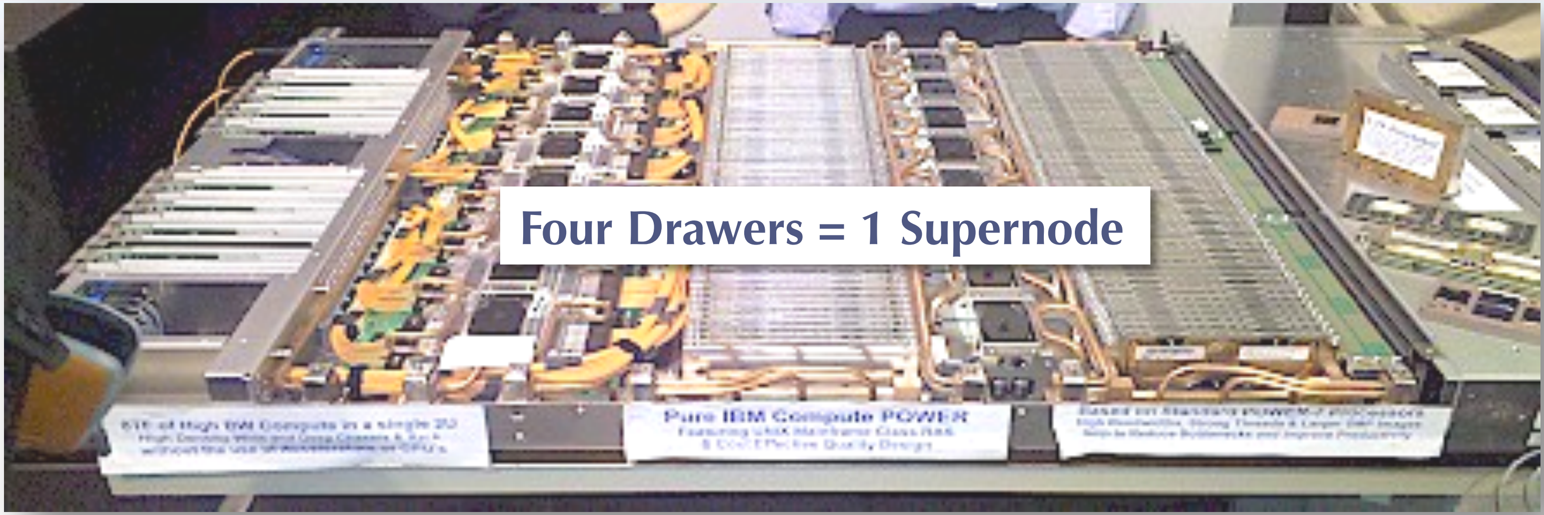




- Quad-chip Module
  - Four Power7 chips
    - 32 cores, 128 threads
    - 1 TF peak performance
    - 128 GBytes memory
      - 512 GB/sec memory bw
  - One Hub Chip
    - 1.128 TB/sec bandwidth
      - 192 GB/s QCM connection
      - 336 GB/s to 7 to other local nodes
      - 240 GB/s to local-remote nodes
      - 320 GB/s to remote nodes
      - 40 GB/s general purpose I/O

Blue Waters Project

IH Server Node



Specifications:

8 Quad-chip Modules
- 8 TFlops (*peak*)

1 TByte of memory
- 4 TB/s memory bw

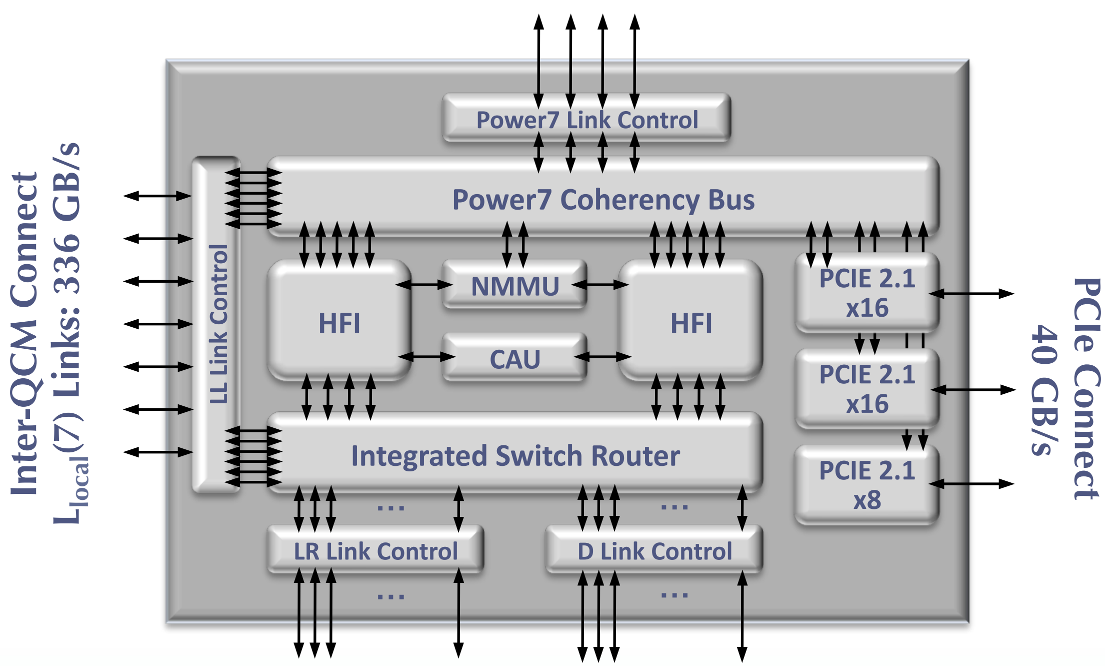
8 Hub Chips
- 9 TB/sec/Hub

Packaging:

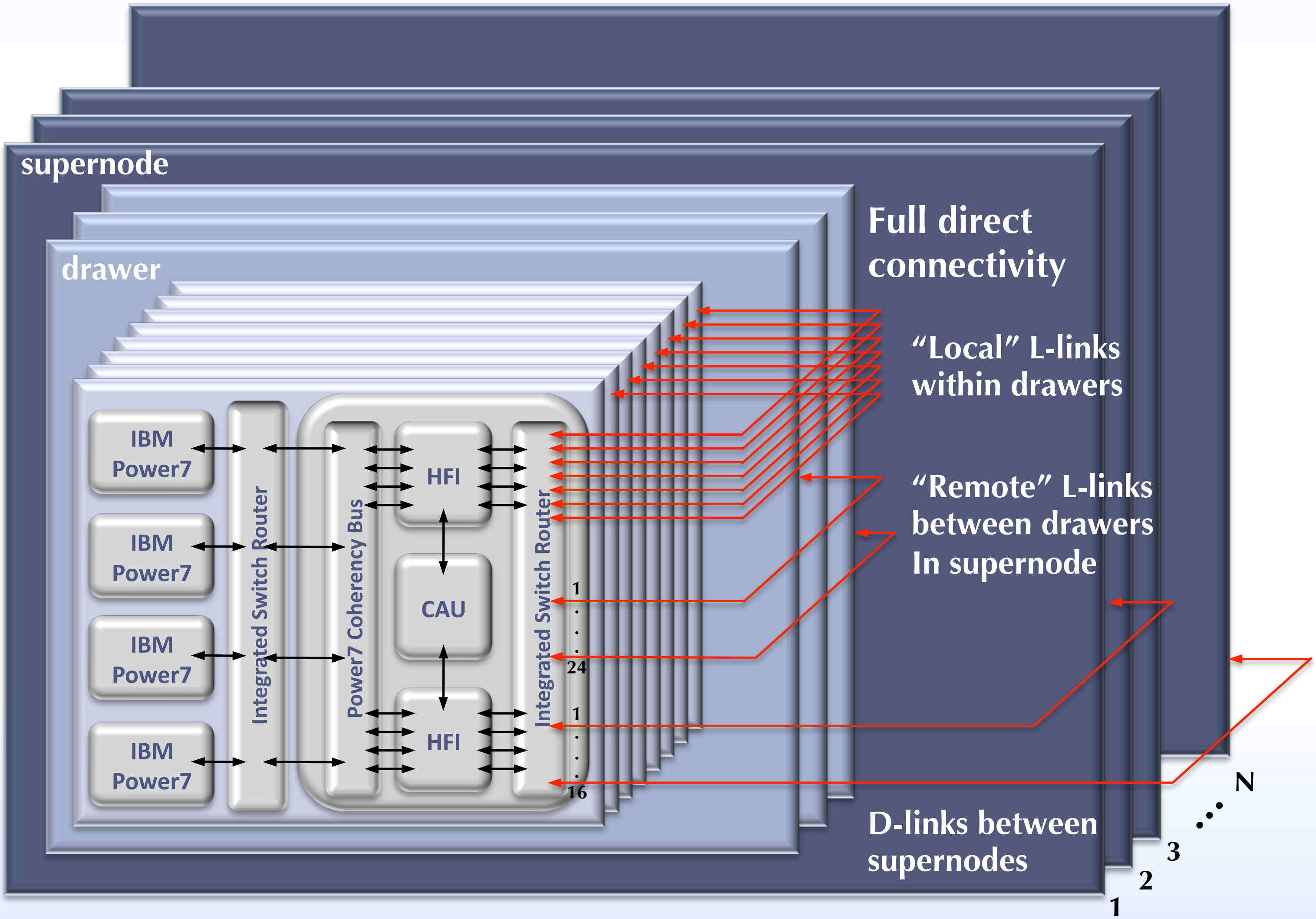
2U Drawer
39"w **x** 72"d
> 300 lbs.

Fully water cooled (QCMs, Hubs, Memory)

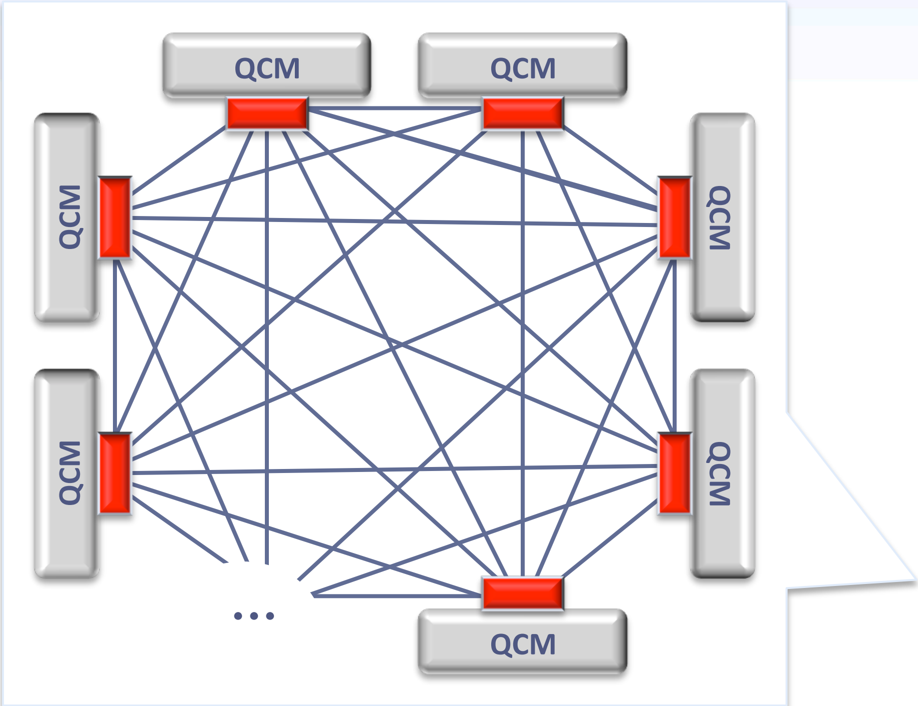
PERCS Hub Chip: It's All About Connections

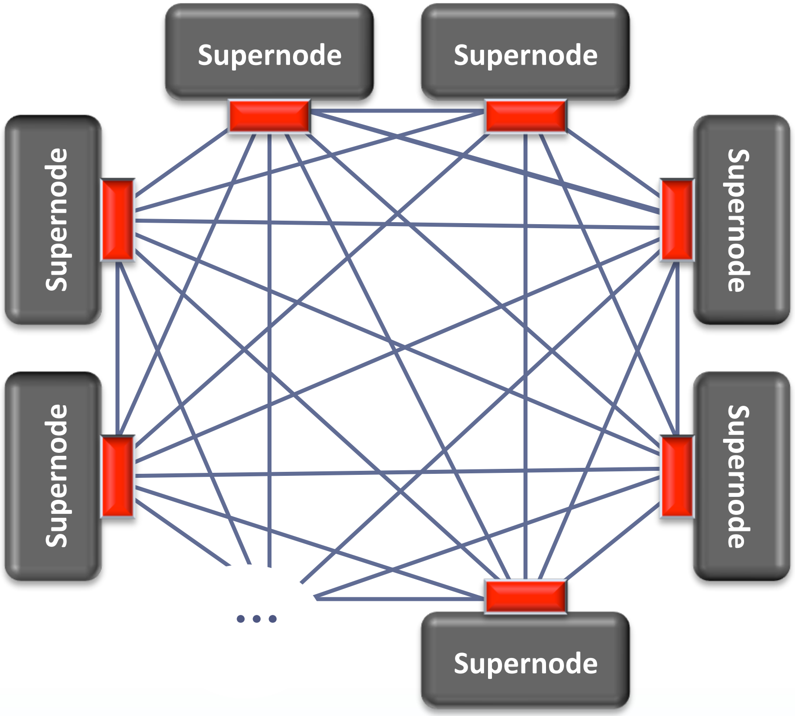


Logical View of Blue Waters Interconnect



Two-level (L, D) Direct-connect Network





Each Supernode = 32 QCMs
(4 Drawers x 8 SMPs/Drawer)

Fully Interconnected with L_{local} and L_{remote} Links

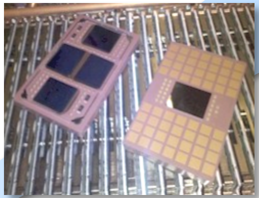
Blue Waters = 320 Supernodes
(40 BBs x 8 SNs/BB)

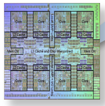
Fully Interconnected with D Links

Blue Waters Project

Building Blue Waters

Blue Waters will be the most powerful computer in the world for scientific research when it comes on line in 2011-2.





Blue Waters

≥10 PF Peak
~1 PF sustained
≥300,000 cores
≥1 PB of memory
>25 PB of disk storage
500 PB of archival storage
≥100 Gbps connectivity

Blue Waters 3-Rack Building Block

32 IH server nodes
- 256 TF (peak)
- 32 TB memory
- 128 TB/s memory bw

4 Storage systems (>500 TB)
10 Tape drive connections

IH Server Node

8 QCM's (256 cores)
- 8 TF (*peak*)

1 TB memory
- 4 TB/s memory bw

8 Hub chips
- 9 TB/s comm bw

Power supplies
PCIe slots

Fully water cooled

Quad-chip Module

4 Power7 chips
- 1 TF (*peak*)

128 GB memory
- 512 GB/s memory bw

Hub Chip

1.128 TB/s comm bw

Power7 Chip

8 cores, 32 threads
L1, L2, L3 cache (32 MB)
Up to 256 GF (peak)
128 Gb/s memory bw

45 nm technology

Blue Waters is built from components that can also be used to build systems with a wide range of capabilities—from deskside to beyond Blue Waters.

Blue Waters and Jaguar Computing Systems

	ORNL		NCSA
System Attribute	**Jaguar (#2)**		**Blue Waters**
Vendor (Model)	Cray (XT5)		IBM (PERCS)
Processor	AMD Opteron		IBM Power7
Peak Performance (PF)	2.3	**~4x**	≳10
Sustained Performance (PF)	?		≳1
Number of Cores/Chip	6	**1⅓x**	8
Number of Processor Cores	224,256	**<1½x**	≳300,000
Amount of Memory (TB)	299	**4x**	≳1,200
Memory Bandwidth (PB/sec)	0.478	**10x**	≳5
Amount of On-line Disk Storage (PB)	10	**>2½x**	≳25
Disk Transfer Rate (TB/sec)	0.24	**>16x**	>4.0

Blue Waters Project

Blue Waters vs. Other Future Systems

System Attribute	NCSA Blue Waters	ANL Mira	RIKEN K-Computer
Vendor (Model)	IBM (PERCS)	IBM (BlueGene)	Fujitsu
Processor	IBM Power7	PowerPC	SPARC64 VIIIfx
Performance			
Peak Performance (PF)	10	10	>10
Sustained Performance (PF)	≳1	?	?
Microprocessor			
Cores/Chip	8	16	8
Peak Performance/Core (GFs)	~32	~12.5	16
Total Processor Cores	>300,000	~750,000	>640,000
Memory			
Main Memory (TB)	1,200	750	1,200
Total Memory Bandwidth (PB/s)	5	2	5
Storage			
Total On-line Disk Storage	>25	70*	
Disk Transfer (TB/sec)	>4.0	0.25–0.5*	
Archival Storage (PB)	*up to* 500	~100**	?
Network			
Interconnect Topology	2-Level Direct	5D Torus	Mesh/Torus
Bisection Bandwidth (TB/s)	~500	30**	>30

Note: All numbers should be regarded as reliable estimates.

Watson and Blue Waters

- **Watson**
  - 90 QCMs
    - 78 TFs (*peak*)
  - 23 TBs memory
    - 256 GB/QCM
  - 4 TB disk storage
- **Blue Waters**
  - 10,000 QCMs
    - 10,000 TFs (*peak*)
  - 1,200 TBs memory
    - 128 GB/QCM
  - >25,000 TB disk storage

Blue Waters Project

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 Project

Machine Room & Subfloor

Machine Room Gallery

Machine Room Subfloor

Blue Waters Project

Great Lakes Consortium for Petascale Computation

Goal: Facilitate the widespread and effective use of petascale computing to address frontier research questions in science, technology and engineering at research, educational and industrial organizations across the region and nation.

Charter Members

Argonne National Laboratory
Fermi National Accelerator Laboratory
Illinois Math and Science Academy
Illinois Wesleyan University
Indiana University*
Iowa State University
Illinois Mathematics and Science Academy
Krell Institute, Inc.
Louisiana State University
Michigan State University*
Northwestern University*
Parkland Community College
Pennsylvania State University*
Purdue University*

The Ohio State University*
Shiloh Community Unit School District #1
Shodor Education Foundation, Inc.
SURA – 60 plus universities
University of Chicago*
University of Illinois at Chicago*
University of Illinois at Urbana-Champaign*
University of Iowa*
University of Michigan*
University of Minnesota*
University of North Carolina–Chapel Hill
University of Wisconsin–Madison*
Wayne City High School

*** CIC universities***

Joint IBM-Illinois Projects. I

- **Computing Systems Software**
  - *Goal*: enhance IBM's HPC software stack
  - Examples
    - Integrated System Management Console
    - Optimized routing algorithms
- **Applications Development**
  - *Goal*: facilitate the development of petascale applications
  - Examples
    - Eclipse-based integrated application development environment
    - Improvement of compilers (xlf, xlc, xlupc, xlcaf)
    - Optimized numerical, communications and I/O libraries
    - Advancement of programming models

Joint IBM-Illinois Projects. II

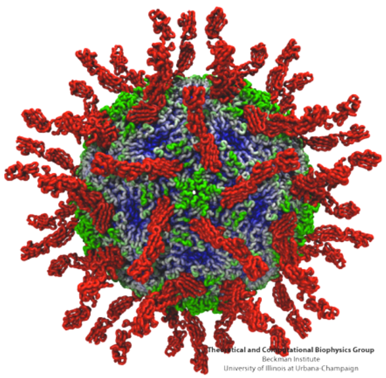
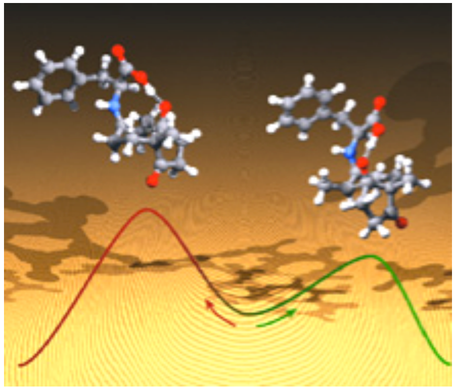
- **Science and Engineering Applications**
  - *Goal*: Optimize performance of petascale applications
  - Examples
    - Optimize application performance on core, chip, drawer, supernode, and system
    - Optimize I/O for data-intensive applications
    - Develop computational models of applications

Science & Engineering Research on Blue Waters

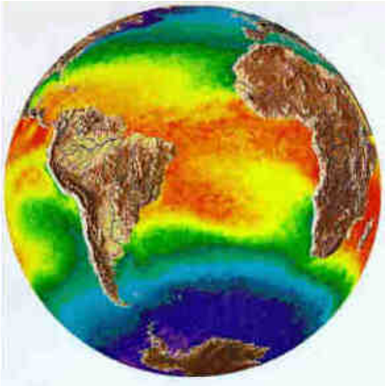
Computational Science and Engineering

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

Molecular Science

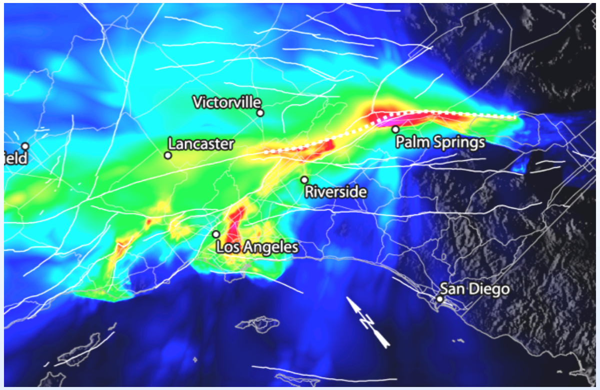



Weather & Climate Forecasting

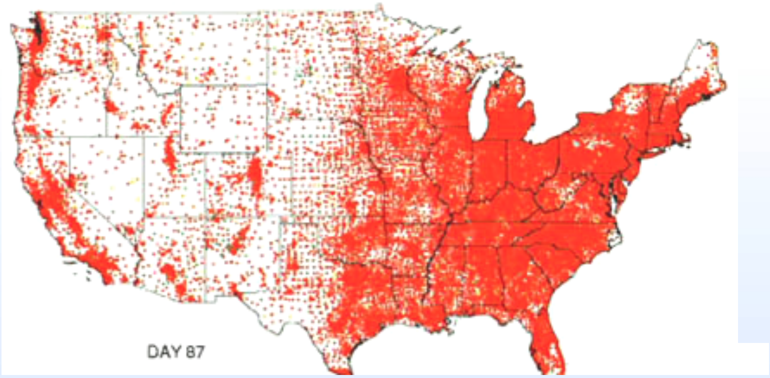



Astronomy

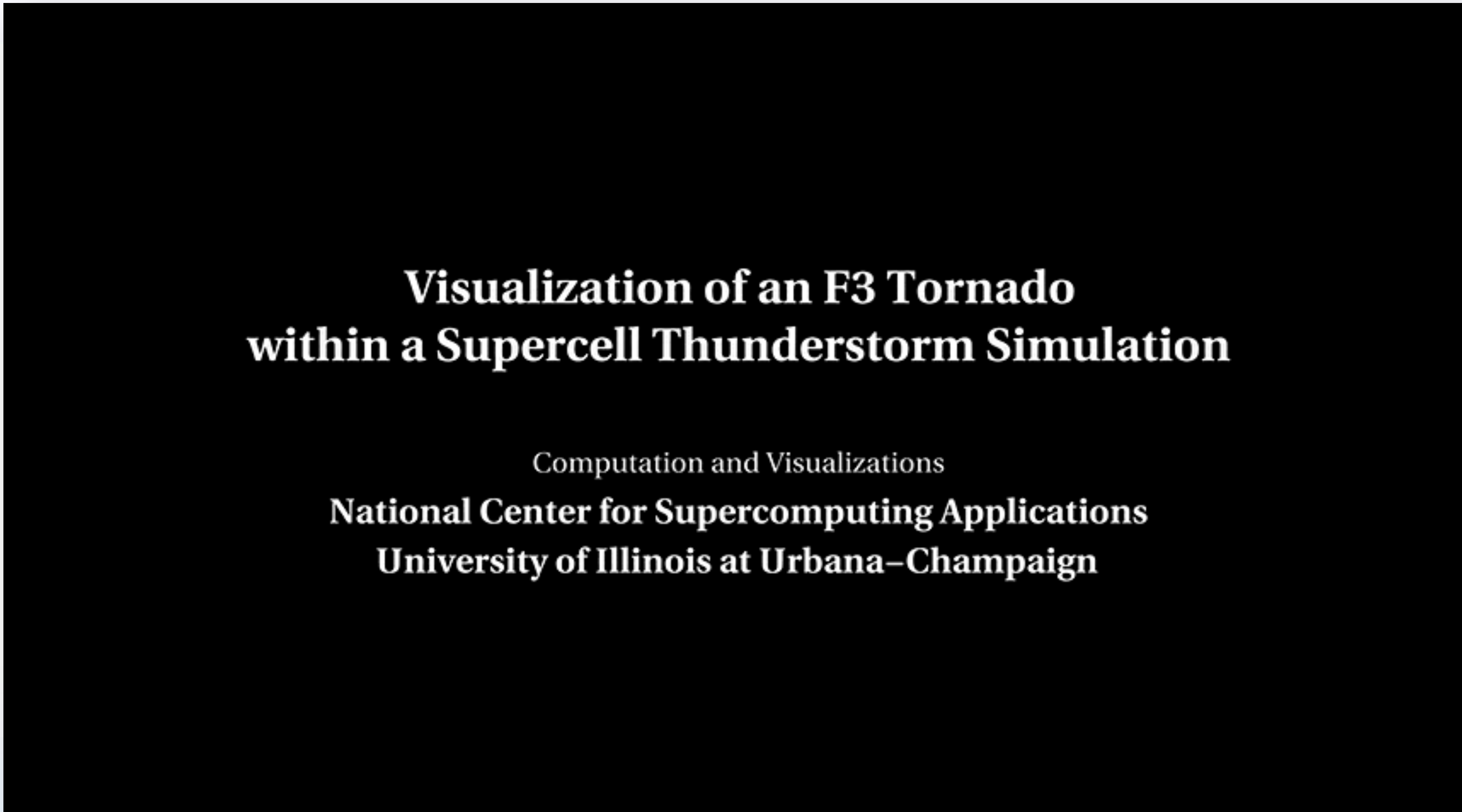
Earth Science



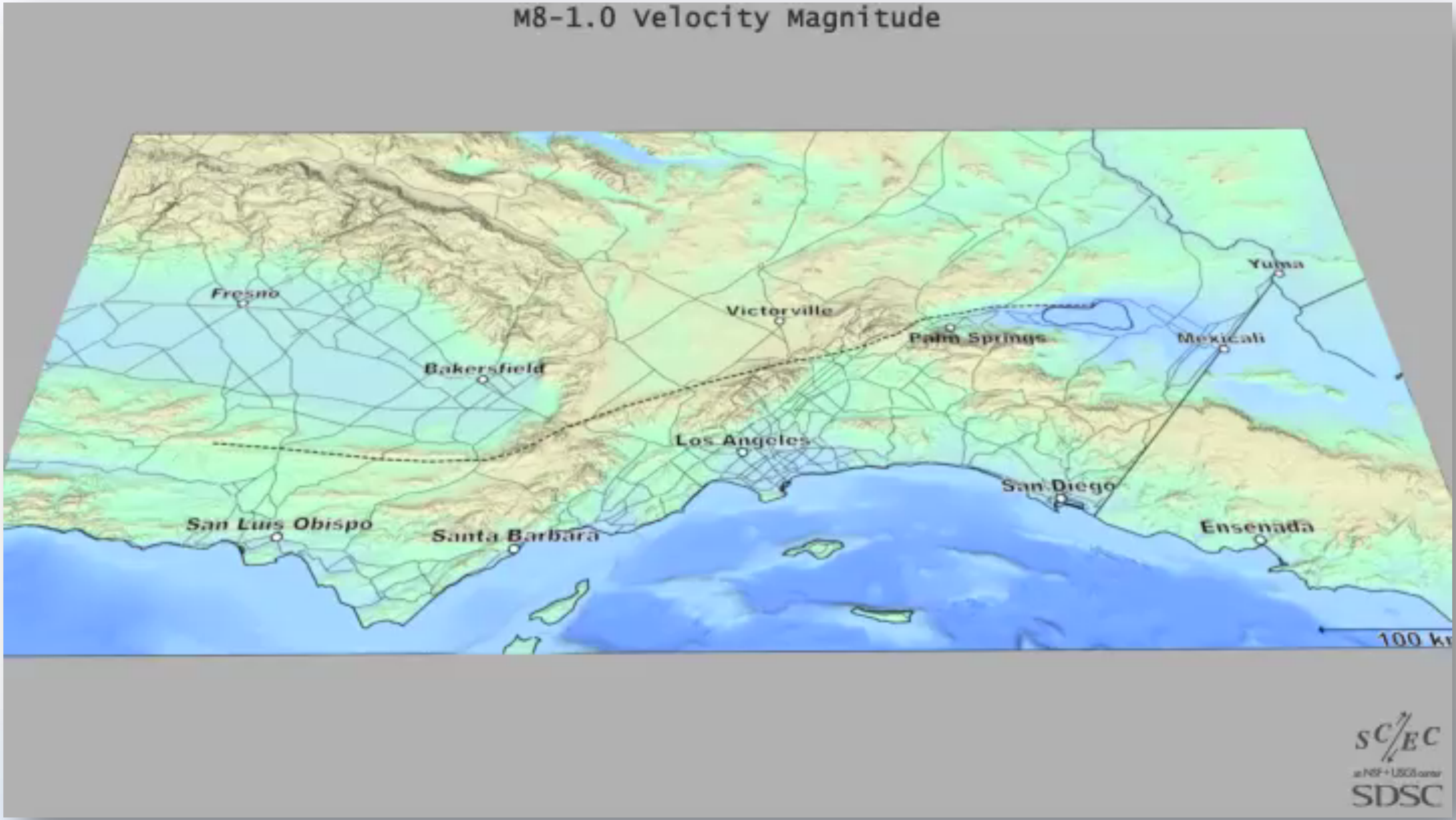
Health



Birth of a Tornado



Propagation of Seismic Waves

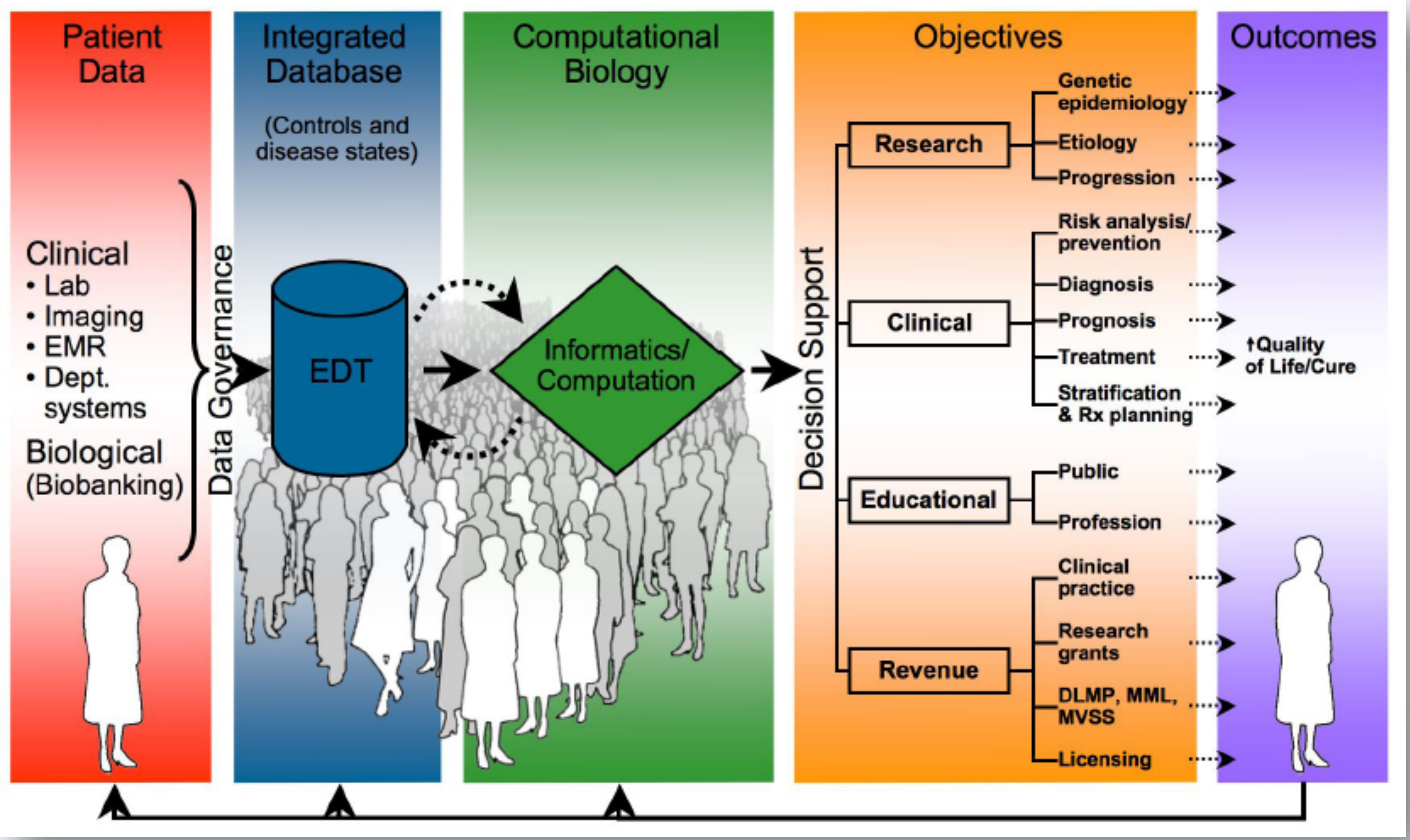


Data-intensive Computing

Large Synoptic Survey Telescope

- **LSST**
  - Located in Chile (El Peñon) with first light in 2013-4
  - 8.4-m Mirror with 3 Gigapixel camera
  - Image available sky every 3 days
- **Science Missions**
  - Nature of dark energy and accelerating universe
  - Comprehensive census of solar system objects, create galactic map
  - Explore transients and variable objects
- **Data Sets**
  - 15–30 terabytes per night (raw)
  - 100–200 petabytes in 10 years

Individualized Medical Informatics



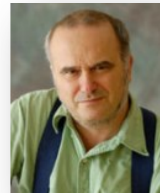
MAYO CLINIC + UNIVERSITY OF ILLINOIS
DIVISION OF BIOMEDICAL SCIENCES

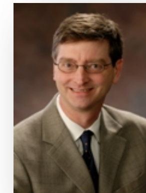
Illinois' Extreme-scale Computing Team

PI & Co-PIs


Dunning


Kramer


Snir


Gropp


Hwu

Task Leads


Beldica
Proj. Mgmt.


Bode
Software


Butler
Storage


Glotzer
Virtual School


Iyer
Reliability


Kale
Apps Simulations


Lathrop
Education


Melchi
Facilities


Giles
Industry


Olson
Proj. Mgmt.


Panoff
UG Education


Semararo
Visualization


Towns
Ops Transition

Questions?