

# TeraGrid TeraGrid and the Path to Petascale

**John Towns**

Chair, TeraGrid Forum

Director, Persistent Infrastructure

National Center for Supercomputing Applications

University of Illinois

[jtowns@ncsa.illinois.edu](mailto:jtowns@ncsa.illinois.edu)



# What is the TeraGrid?

- **World's largest open scientific discovery infrastructure**
  - supported by US National Science Foundation
  - extremely user-driven
    - MPI jobs, ssh or grid (GRAM) access, etc.
- **An instrument that delivers high-end resources and services**
  - a computational facility: over a Pflop/s in parallel computing capability
    - will grow to > 1.75 Pflop/s in 2009
  - high performance networks
  - a data storage and management facility: over 30 PB of storage, over 100 scientific data collections
  - visualization systems, Science Gateways, User Portal
- **A service: help desk and consulting, Advanced User Support (AUS), education and training events and resources**
- **Something you can use without financial cost**
  - allocated to US researchers and their collaborators through national peer-review process
    - generally, review of computing, not science



# Our Vision of TeraGrid

- **Three part mission:**
  - support the most advanced computational science in multiple domains
    - address key challenges prioritized by users
  - empower new communities of users
    - partner with science community leaders - "Science Gateways"
  - provide resources and services that can be extended to a broader cyberinfrastructure
    - partner with campuses and facilities
- **TeraGrid is...**
  - an advanced, nationally distributed, open cyberinfrastructure comprised of supercomputing, storage, and visualization systems, data collections, and science gateways, integrated by software services and high bandwidth networks, coordinated through common policies and operations, and supported by computing and technology experts, that enables and supports leading-edge scientific discovery and promotes science and technology education
  - a complex collaboration of over a dozen organizations and NSF awards working together to provide collective services that go beyond what can be provided by individual institutions



# TeraGrid: greater than the sum of its parts...

- Single unified allocations process
- Single point of contact for problem reporting and tracking
  - especially useful for problems between systems
- Simplified access to high end resources for science and engineering
  - single sign-on
  - coordinated software environments
  - uniform access to heterogeneous resources to solve a single scientific problem
  - simplified data movement
- Expertise in building national computing and data resources
- Leveraging extensive resources, expertise, R&D, and EOT
  - leveraging other activities at participant sites
  - learning from each other improves expertise of all TG staff
- Leadership in cyberinfrastructure development, deployment and support

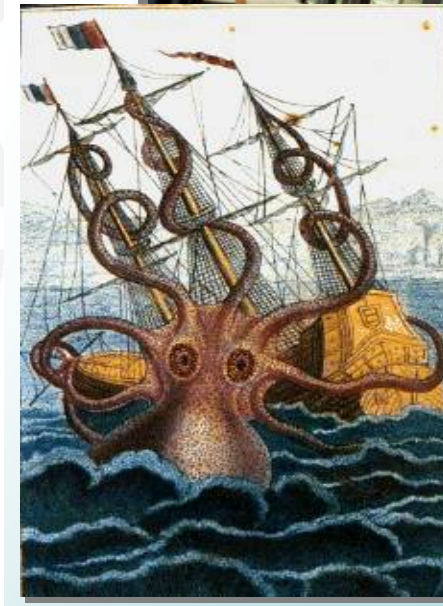


– demonstrating enablement of science not possible without the TeraGrid-coordinated human and technological resources



# Diversity of Resources (not exhaustive)

- **Very Powerful Tightly Coupled Distributed Memory**
  - Ranger (TACC): Sun Constellation, 62,976 cores, 579 Tflop/s, 123 TB RAM
  - Kraken (NICS): Cray XT5, 66,048 cores, 608 Tflop/s, > 1 Pflop/s in 2009
- **Shared Memory**
  - Cobalt (NCSA): Altix, 8 Tflop/s, 3 TB shared memory
  - Pople (PSC): Altix, 5 Tflop/s, 1.5 TB shared memory
- **Clusters with Infiniband**
  - Abe (NCSA): 90 Tflop/s
  - Lonestar (TACC): 61 Tflop/s
  - QueenBee (LONI): 51 Tflop/s
- **Condor Pool (Loosely Coupled)**
  - Purdue- up to 22,000 CPUs
- **Visualization Resources**
  - TeraDRE (Purdue): 48 node nVIDIA GPUs
  - Spur (TACC): 32 nVIDIA GPUs
- **Storage Resources**
  - GPFS-WAN (SDSC)
  - Lustre-WAN (IU)
  - various archival resources



# Resources to come...

- **Track 2c @ PSC**
  - large shared memory system in 2010
- **Track 2d being competed**
  - data-intensive HPC system
  - experimental HPC system
  - pool of loosely coupled, high throughput resources
  - experimental, high-performance grid test bed
- **eXtreme Digital (XD) High-Performance Remote Visualization and Data Analysis Services**
  - service and possibly resources; up to 2 awards (?)
- **Blue Waters (Track 1) @ NCSA:**
  - 1 Pflop/s sustained on serious applications in 2011
- **Unsolicited proposal for archival storage enhancements pending**

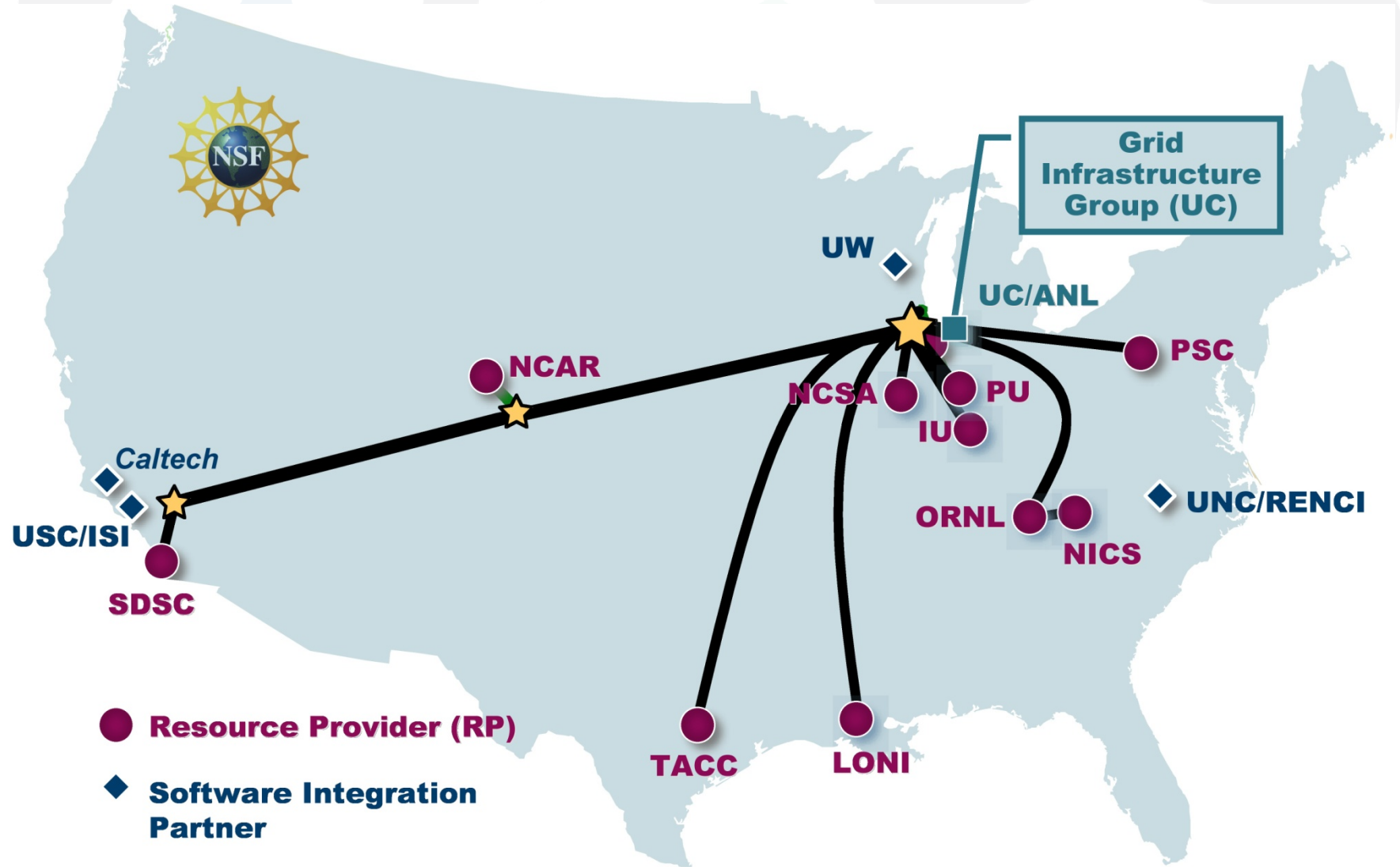


# How is TeraGrid Organized?

- TG is set up like a large cooperative research group
  - evolved from many years of collaborative arrangements between the centers
  - still evolving!
- Federation of 12 awards
  - Resource Providers (RPs)
  - Grid Infrastructure Group (GIG)
- Strategically lead by the TeraGrid Forum
  - made up of the PI's from each RP and the GIG
  - led by the TG Forum Chair, who is responsible for coordinating the group (elected position)
    - John Towns – TG Forum Chair
  - responsible for the strategic decision making that affects the collaboration
- Centrally coordinated by the GIG



# TeraGrid Participants



# Who are the Players?

## • GIG Management

- GIG Director: Matthew Heinzl
- GIG Director of Science: Dan Katz
- Area Directors:
  - Software Integration: Lee Liming/J.P. Navarro
  - Gateways: Nancy Wilkins-Diehr
  - User Services: Sergiu Sanielevici
  - Advanced User Support: Amit Majumdar
  - Data and Visualization: Kelly Gaither
  - Network, Ops, and Security: Von Welch
  - EOT: Scott Lathrop
  - Project Management: Tim Cockerill
  - User Facing Projects and Core Services: Dave Hart

## • TeraGrid Forum

- TG Forum Chair: John Towns
- Membership:
  - PSC: Ralph Roskies
  - NICS: Phil Andrews
  - ORNL: John Cobb
  - Indiana: Craig Stewart
  - Purdue: Carol Song
  - U Chicago/ANL: Mike Papka
  - NCSA: John Towns
  - LONI: Dan Katz
  - TACC: Jay Boisseau
  - NCAR: Rich Loft
  - SDSC: Richard Moore
  - GIG: Matt Heinzl

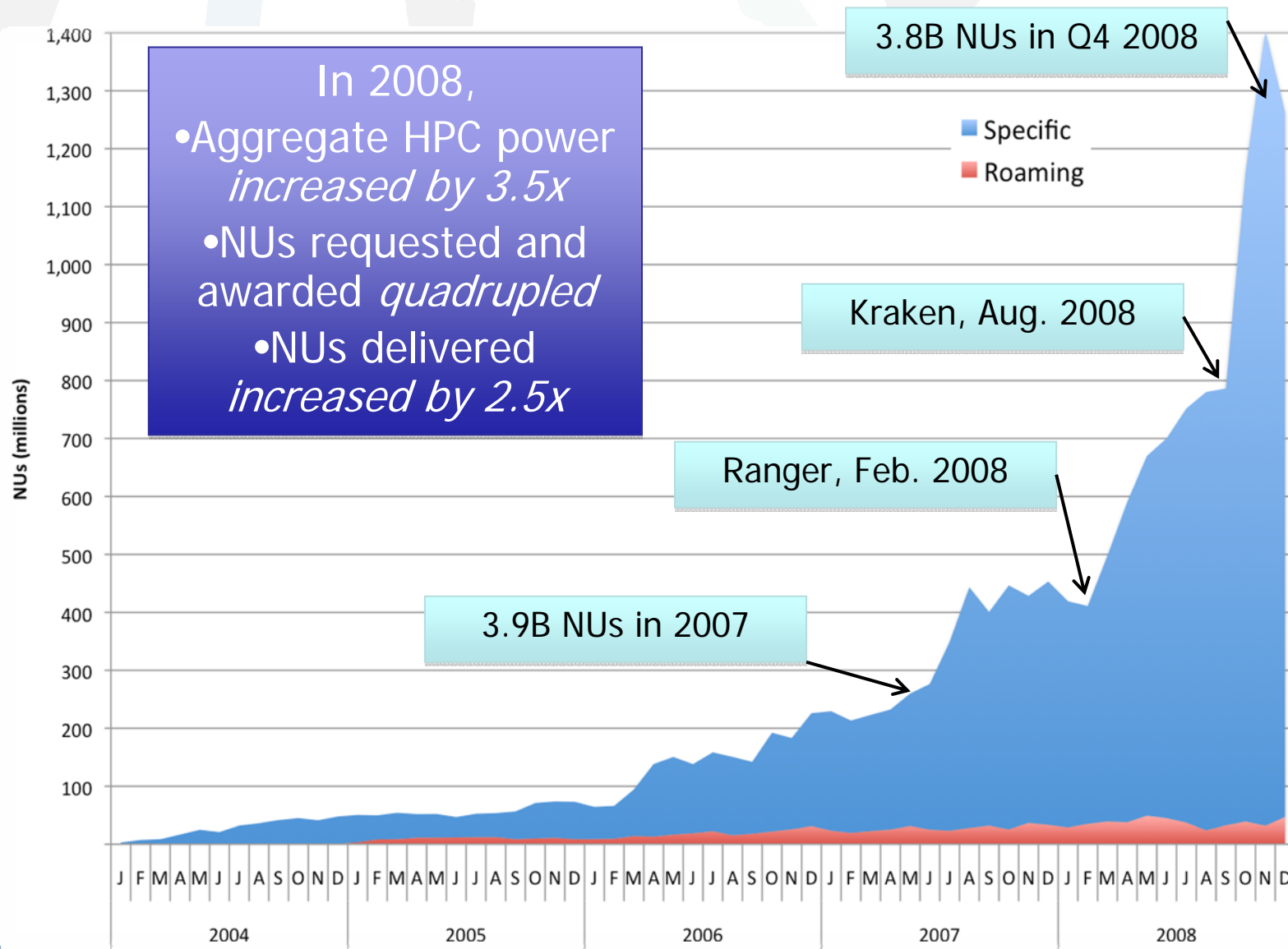


# Allocations Process

- **National peer-review process**
  - allocates computational, data, and visualization resources
  - makes recommendations on allocation of advanced direct support services
- **Managed by TeraGrid**
  - GIG and RP Participants in reviews
  - CORE Services award to manage shared responsibilities
    - TACC: Meeting coordination
    - SDSC: TG Central DB
    - NCSA: POPS, TG Allocations group
- **Currently awarding >10B Normalized Units of resources annually**

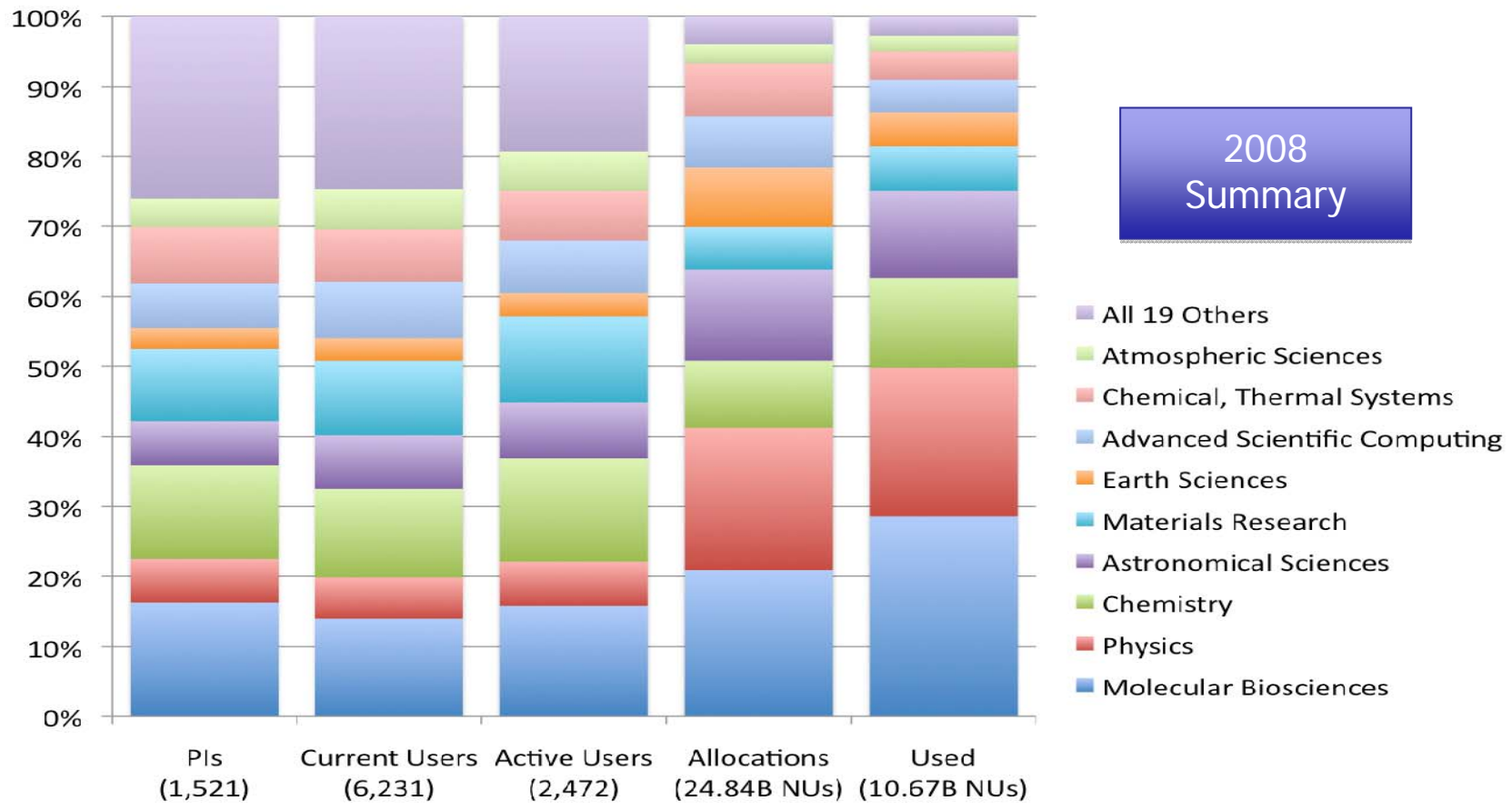


# TeraGrid HPC Usage, 2008



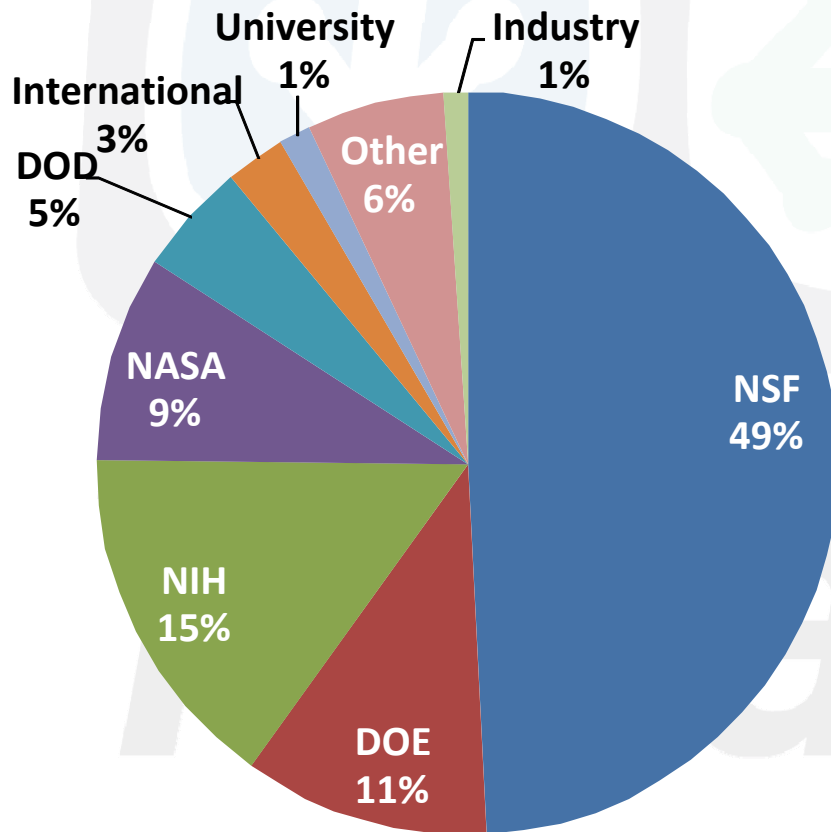
# TeraGrid Use by Discipline, 2008

~2,500 users charged jobs in 2008  
 Representing 332 institutions, 48 states + PR,DC



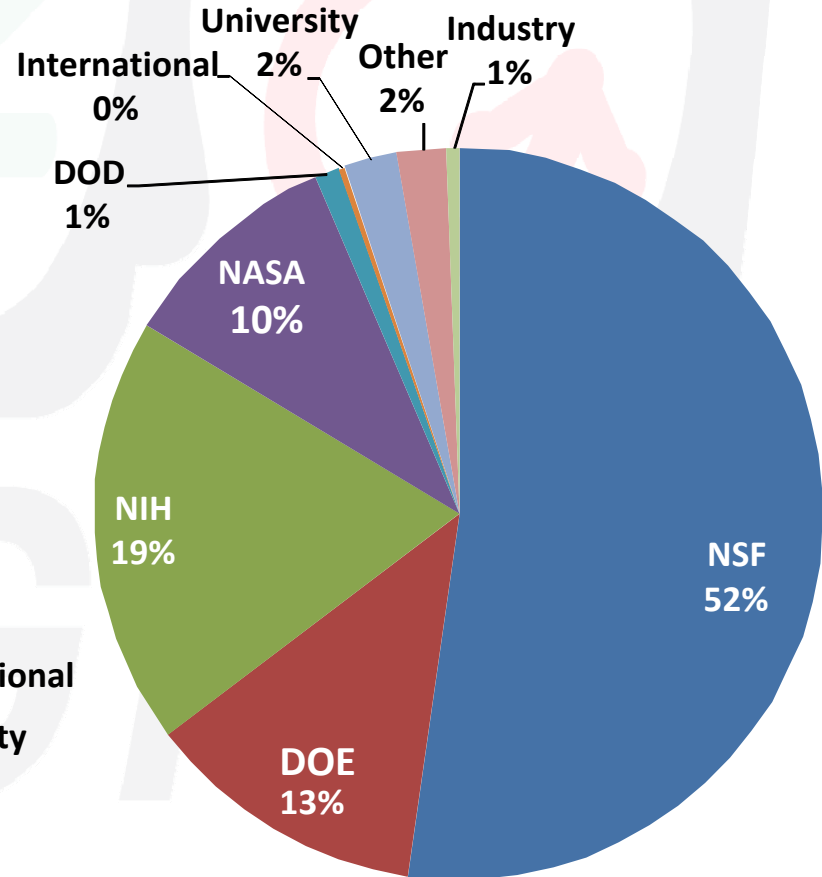
# Impacting Many Agencies

Supported Research Funding by Agency



*\$91.5M in Funded Research Supported*

Resource Usage by Agency



*10B NUs Delivered*

- NSF
- DOE
- NIH
- NASA
- DOD
- International
- University
- Other
- Industry



# Geosciences (SCEC)

- Goal is understanding earthquakes and to mitigate risks of loss of life and property damage.
- Spans the gamut from largest simulations to midsize jobs to huge number of small jobs
- For largest runs (Cybershake), where they examine high frequency modes (short wavelength, so higher resolution) of particular interest to civil engineers, need large distributed memory runs using the Track 2 machines at TACC, NICS; 2,000-64,000 cores of Ranger, Kraken.
- To improve the velocity model that goes into the large simulations, need mid-range core counts jobs doing full 3-D tomography (Tera3D); DTF and other clusters (e.g. Abe); Need large data available on disk (100 TB)



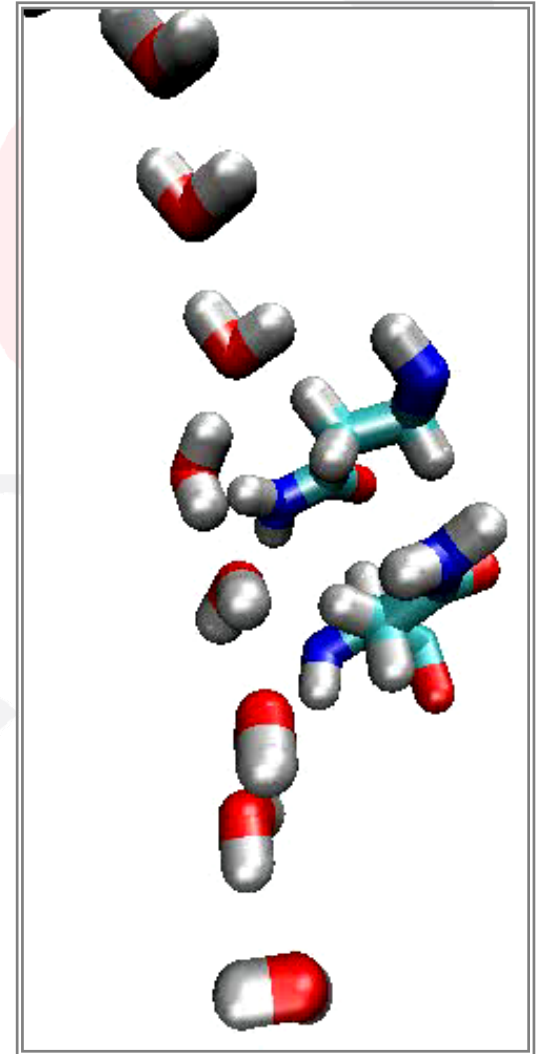
Excellent example of coordinated ASTA support- CUI (SDSC) and Urbanic (PSC) interface with consultants at NICS, TACC, & NCSA to smooth migration of code. Improved performance 4x.

Output is large data sets stored at NCSA, or SDSC's GPFS, IRODS. Moving to DOE machine at Argonne. TG provided help with essential data transfer.

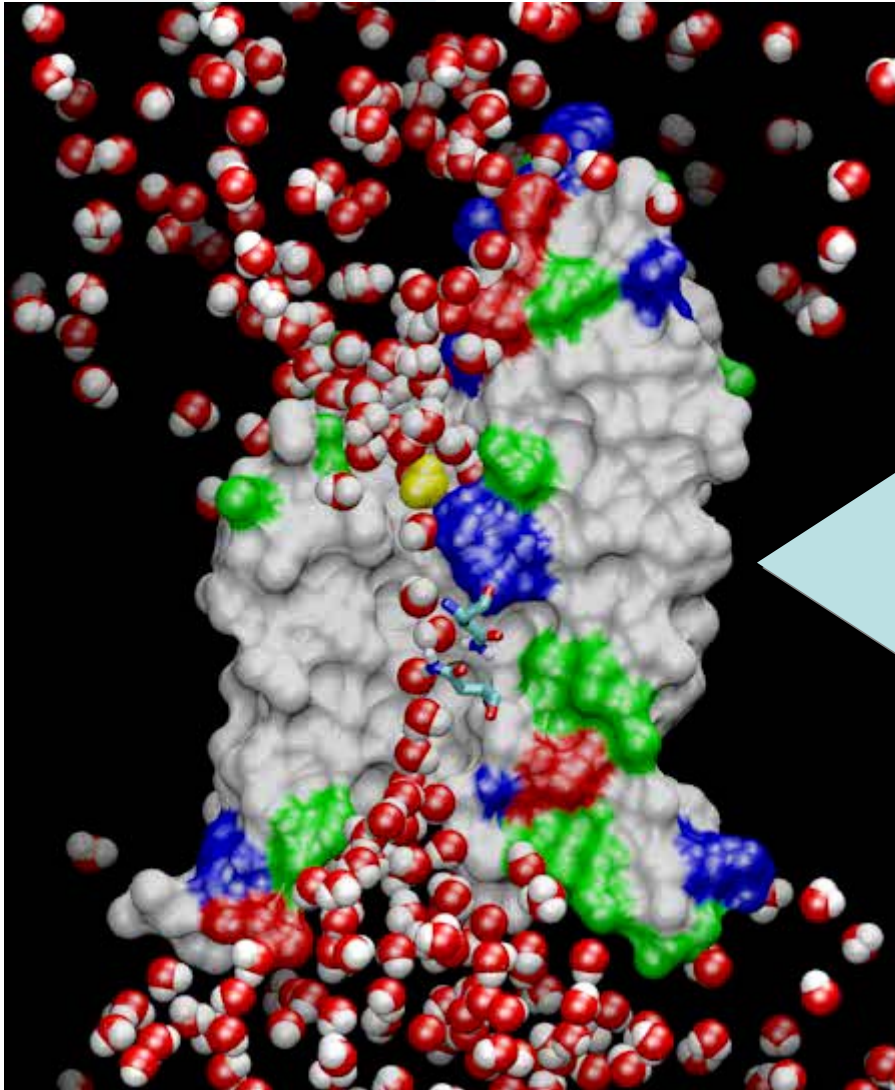


# Aquaporins - Schulten group, UIUC

- Aquaporins are proteins which conduct large volumes of water through cell walls while filtering out charged particles like hydrogen ions (protons).
- Start with known crystal structure, simulate over 100,000 atoms, using NAMD
- Water moves through aquaporin channels in single file. Oxygen leads the way in. At the most constricted point of channel, water molecule flips. Protons can't do this.



# Aquaporin Mechanism



Animation pointed to by 2003 Nobel chemistry prize announcement for structure of aquaporins (Peter Agre)

The simulation helped explain how the structure led to the function



# Where is TeraGrid Going?

- **Current Program**
  - nominal end date for TeraGrid is March 2010
- **TeraGrid Phase III: eXtreme Digital Resources for Science and Engineering (XD)**
  - follow-on program for TeraGrid
  - four Integrating Services
    - Coordination and Management Service (CMS)
    - Technology Audit and Insertion Service (TAIS)
    - Advanced User Support Service (AUSS)
    - Training, Education and Outreach Service (TEOS)
  - original planned start date on April 2010
    - CMS, AUSS and TEOS deferred one year



# TeraGrid → TeraGrid Extension (?) → XD Transition Planning

- All current TeraGrid activity areas have effort reserved for TeraGrid → XD transition effort as appropriate
  - transition issues exist for nearly all areas
    - effort ear-marked to support transition issues
- Start of XD for CMS/AUSS/TEOS deferred for one year (1 April 2011)
  - induced TeraGrid Extension Proposal
    - 12-month funding to support most GIG functions and some non-Track 2 RP resources
  - uncertainty in sequence of events
  - still need to address many changes in TeraGrid going into presumed Extension Period
    - many resources exit TeraGrid
- Program Year 5 planning process will *likely* need to address:
  - TeraGrid Extension following PY5
  - necessarily include transition to XD in extension period

