The Challenge of using Petascale Computing for Research

Steve Meacham

National Science Foundation

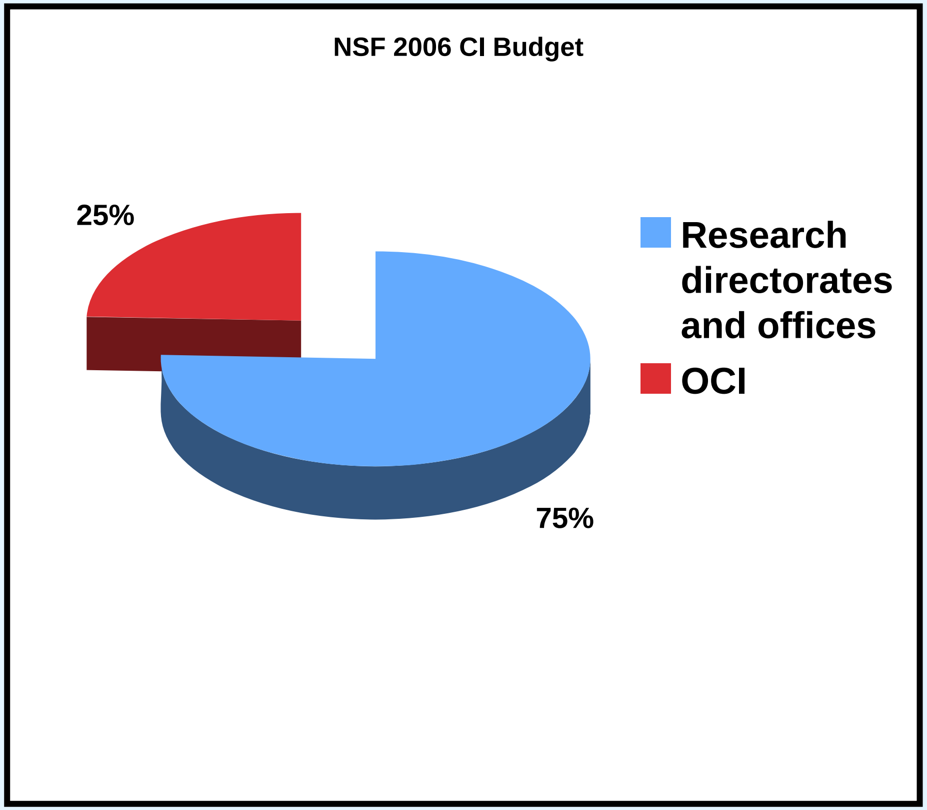
Geosciences High-Performance Computing Workshop

September 25, 2006

Outline

- OCI HPC Investments
- TeraGrid
- Why petascale is a challenge

NSF CI Budgets



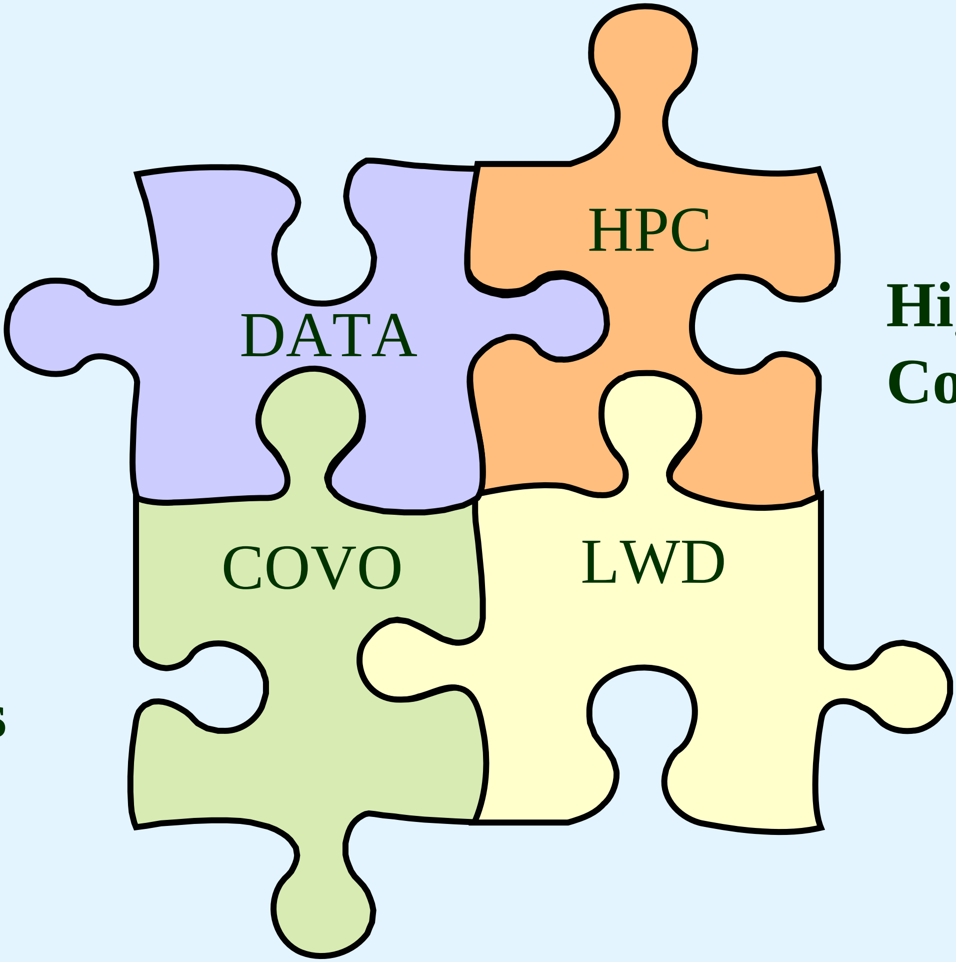
Directorates/Research Offices
-discipline-specific CI, e.g. model development and data collections

Office of Cyberinfrastructure (OCI)
- development & deployment of CI shared by multiple disciplines

OCI Portfolio

Data, Data Analysis, and Visualization

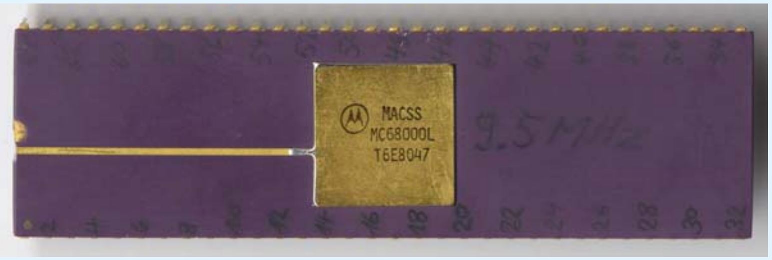
Enhancing Virtual Organizations



High-Performance Computing

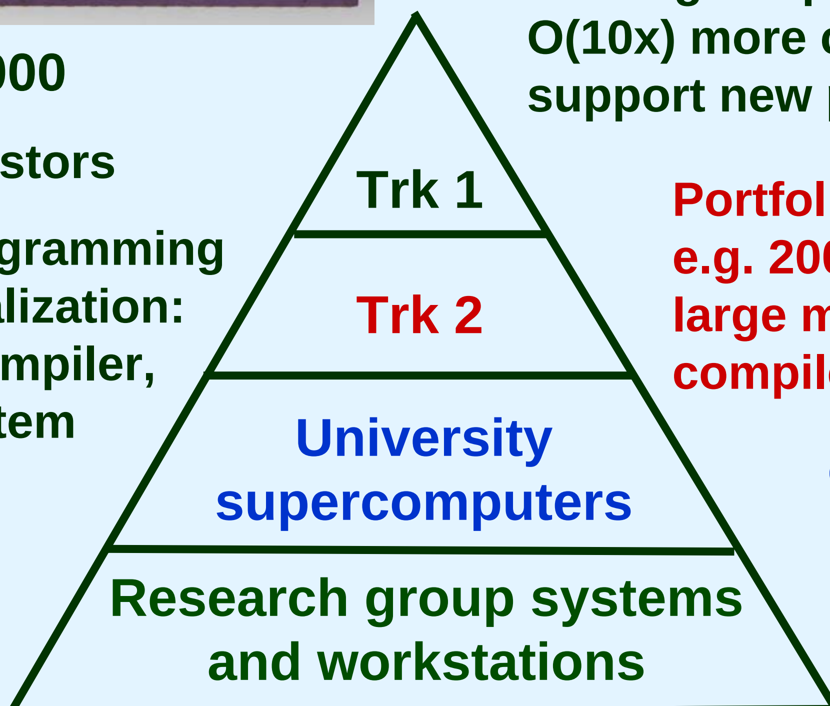
Learning and Workforce Development

HPC spectrum for research



Motorola 68000

- 70000 transistors
- simplify programming through virtualization: assembler, compiler, operating system



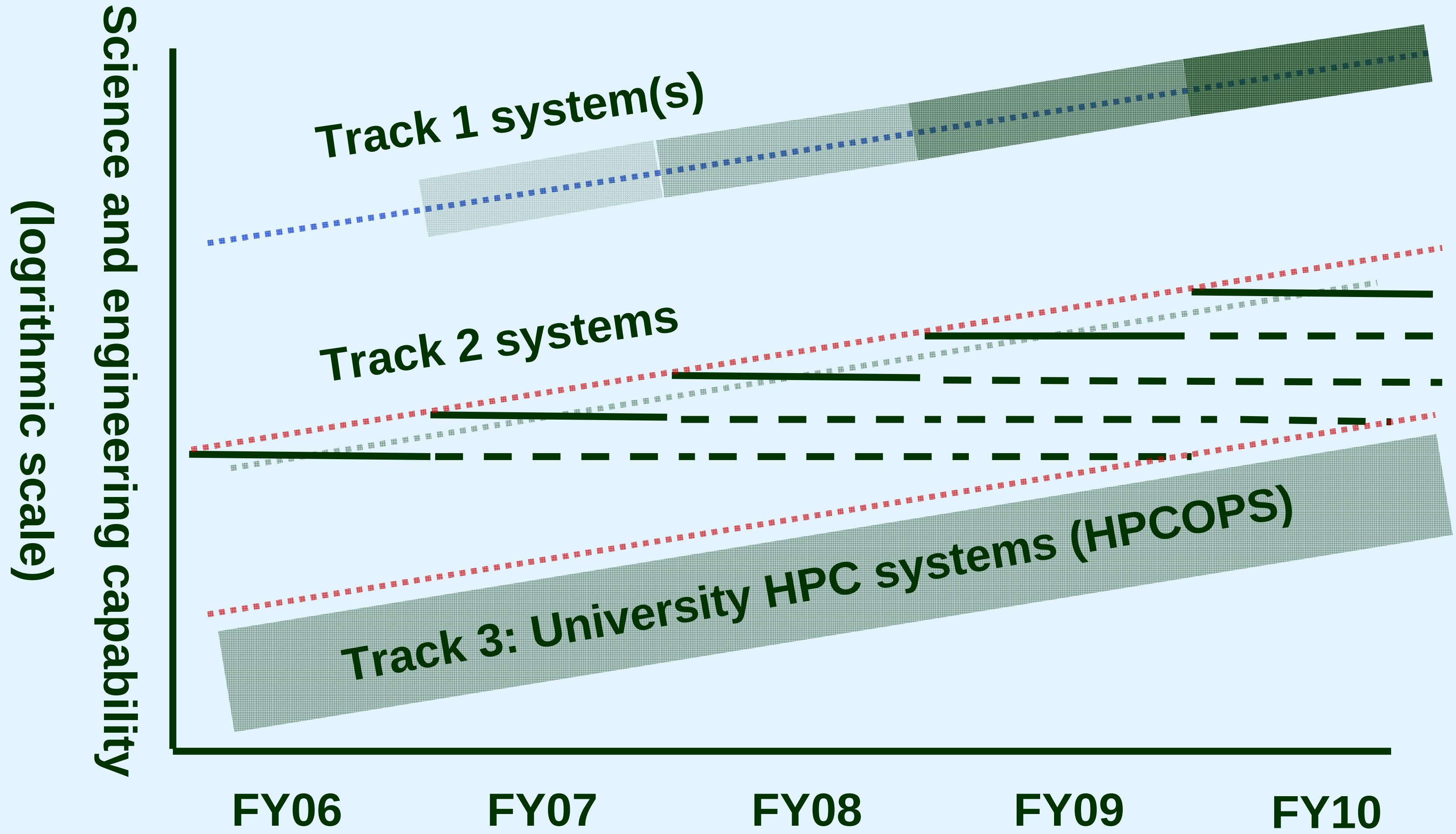
5 years out - capable of sustaining PF/s on range of problems - lots of memory - O(10x) more cores - new system SW - support new programming models

Portfolio of large, powerful systems - e.g. 2007: > 400 TF/s; > 50K cores; large memory - support PGAS compilers

O(1K - 10K) cores

Multi-core

Acquisition Strategy



TeraGrid: an integrating infrastructure

TeraGrid

Offers:

- Common user environments
- Pooled community support expertise
- Targeted consulting services (ASTA)
- Science gateways
- A portfolio of architectures

Exploring:

- A security infrastructure that uses campus authentication systems
- A lightweight, service-based approach to enable campus grids to federate with TeraGrid

TeraGrid

Aims to simplify use of HPC and data through virtualization:

- Single login & TeraGrid User Portal
- Global WAN filesystems
- TeraGrid-wide resource discovery
- Meta-scheduler
- Scientific workflow orchestration
- Science gateways

and productivity tools for large computations

- High-bandwidth I/O between storage and computation
- Remote visualization engines and software
- Analysis tools for very large datasets
- Specialized consulting & training in petascale techniques

Science Gateways

- Specific examples of Virtual Organizations
- Built to serve communities of practice by bring together a variety of resources in a customized portal
- Examples include:
  - NanoHub
  - NEES
  - LEAD
  - SCEC Earthworks Project
  - NVO
- http://www.teragrid.org/programs/sci_gateways/

Challenges to HPC use

- **Trend to large numbers of cores and threads - how to use effectively?**
  - E.g. BG/L at LLNL: 367 TF/s, > 130,000 cores
  - E.g. 2007 Cray XT at ORNL: > 250 TF/s, > 25,000 cores
  - E.g. 2007 Track 2 at TACC: > 400 TF/s, > 50,000 cores
  - Even at workstation-level see dual-core arch. with multiple FP pipelines and processor vendors plan to continue trend
- **How to fully exploit parallelism?**
  - Modern systems have multiple levels with complex hierarchies of latencies and communications bandwidths. How to design tunable algorithms to map to different hierarchies to increase scaling and portability?
- **I/O management** - highly parallel to achieve bandwidth
- **Fault tolerance** - joint effort of system software and applications
- **Hybrid systems**
  - E.g. LANL's RoadRunner (Opteron + Cell BE)

Examples of codes running at scale

- Several codes show scaling on BG/L to 16K cores
  - E.g. HOMME (atmospheric dynamics); POP (ocean dynamics)
  - E.g. Variety of chemistry and materials science codes
  - E.g. DoD fluid codes
- Expect one class of use to be large numbers of replicates (ensembles, parameter searches, optimization, …)
  - BLAST, EnKF
- But takes dedicated effort: DoD and DoE are making use of new programming paradigms, e.g. PGAS compilers, and using teams of physical scientists, computational mathematicians and computer scientists to develop next-generation codes
  - At NSF, see focus on petascale software development in physics, chemistry, materials science, biology, engineering
- Provides optimism that there are a number of areas that will benefit from the new HPC ecosystem

Investments to help the research community get the most out of modern HPC systems

- **DoE SciDAC-2 (Scientific Discovery through Advanced Computing)**
  - **30 projects; $60M annually**
  - **17 Science Application Projects ($26.1M):** groundwater transport, computational biology, fusion, climate (Drake, Randall), turbulence, materials science, chemistry, quantum chromodynamics
  - **9 Centers for Enabling Technologies ($24.3M):** focus on algorithms and techniques for enabling petascale science
  - **4 SciDAC Institutes ($8.2M):** help a broad range of researchers prepare their applications to take advantage of the increasing supercomputing capabilities and foster the next generation of computational scientists
- **DARPA**
  - **HPCS (High-Productivity Computing Systems):**
    - Petascale hardware for the next decade
    - Improved system software and program development tools

Investments to help the research community get the most out of modern HPC systems

- **NSF**
  - **HECURA (High-End Computing University Research Activity):**
    - **FY06: $10M - I/O, filesystems, storage, security**
    - **FY05: $7M - compilers, debugging tools, schedulers etc - w/ DARPA**
  - **Software Development for Cyberinfrastructure: includes a track for improving HPC system software and program development tools**
  - **ENG & BIO - Funding HPC training programs at SDSC**
  - **OCI+MPS+ENG - Developing solicitation to provide funding for groups developing codes to solve science and engineering problems on petascale systems**

Questions facing computational research communities

- How to prioritize investments in different types of cyberinfrastructure
  - HPC
  - Data collections
  - Science Gateways/Virtual Organizations
  - CI to support next-generation observing systems
- Within HPC investments, what is the appropriate balance between hardware, software development, and user support?
- What part of the HPC investment portfolio is best made in collaboration with other disciplines, and what aspects need geo-specific investments?
- What types of support do researchers need to help them move from classical programming models to new programming models?

Thank you.