Computer Science Research

Fred Johnson
FJohnson@er.doe.gov
301-903-3601

Contribution of Program Element to Overall Office of Science Strategic Goals

- **Research, Development, Testing, Evaluation of high performance software infrastructure to:**
  - Support the effective use of HPC systems for scientific applications
  - Extract knowledge from petabytes of simulation and experimental data

Systems Software

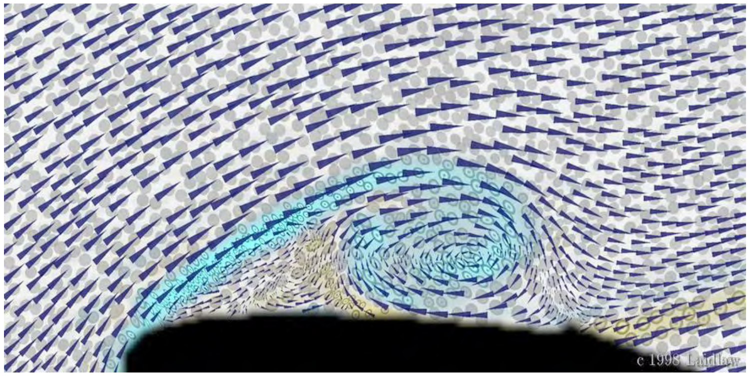
- **Challenge – HPC for Science is (still after fifteen years!)**
  - Hard to use
  - Inefficient
  - Fragile
  - An unimportant vendor market
- **Vision**
  - A comprehensive, integrated software environment which enables the effective application of high performance systems to critical DOE problems
- Goal– Radical Improvement in
  - Application Performance
  - Ease of Use
  - Time to Solution

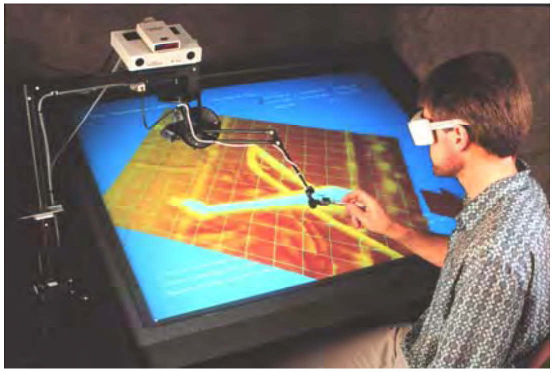
System Admin	Software Development	Scientific Applications
Res. Mgt	Framework	PSEs
Scheduler	Compilers	Viz/Data
Chkpt/Rstrt	Debuggers	Math Libs
File Sys	Perf Tools	Runtme
User Space Runtime Support		
OS Kernel		OS Bypass
Node and System Hardware Arch		

HPC System Elements

Data Management and Visualization

As R.W. Hamming noted, "the purpose of computing is insight, not numbers"





New techniques are needed to glean maximum insight from simulations and experiments which generate petabytes of data.

Program Components

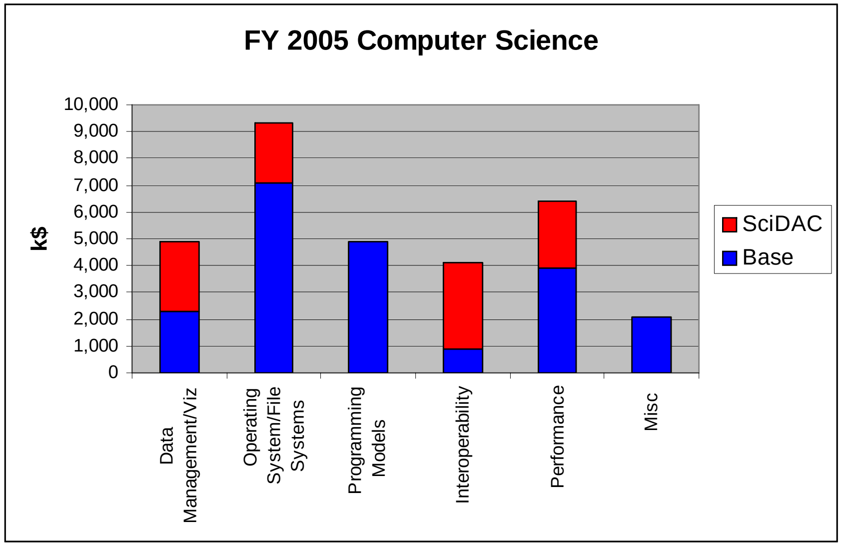
- **Base Program**
  - Evolutionary and revolutionary software methodologies for future generations of HPC architectures

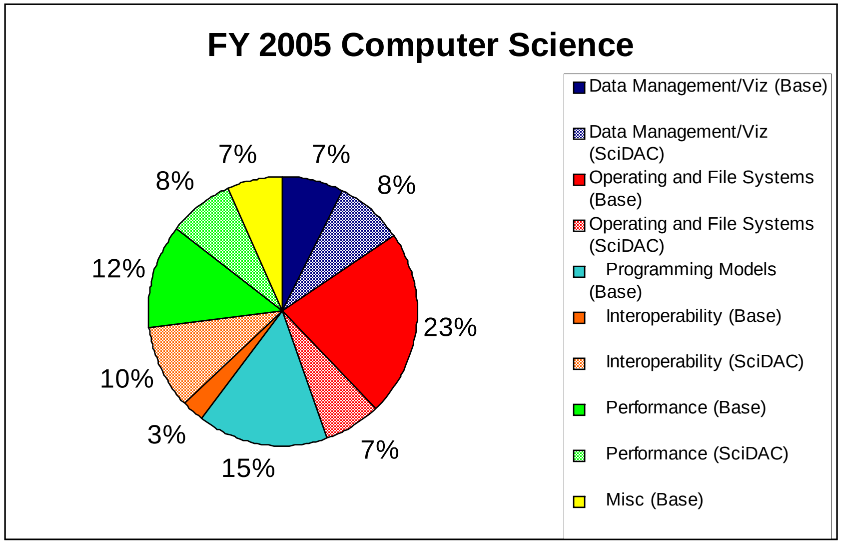
- **SciDAC Integrated Software Infrastructure Centers**
  - Enable effective application of current terascale and emerging petascale architectures to SciDAC applications through focused research and partnerships

Computer Science Budget Breakdown

Advanced Scientific Computing Research Program

FY 2005: Base $21M, SciDAC $11M





FY 2005: Labs 68%, Universities 32%

Major Accomplishments

- **PVM** – the first widely successful model for parallel computing
- **MPI** – the lingua franca of today's parallel computing
- **MPICH** – the open source version of MPI that is the basis for all vendor adaptations
- **Global Arrays** – the distributed shared memory programming model that is at the core of NWChem, the motivating application for SciDAC
- **CTSS** – the first interactive operating system for high performance computers
- SUNMOS/Puma/Cougar – the most successful high performance parallel operating system
- PVFS2/ROMIO – Scalable parallel file system and I/O runtime support

Strengths

- **Particular leadership**
  - Performance modeling, evaluation and optimization
  - Parallel programming models, esp MPI, Global Arrays and Co-Array Fortran
  - Parallel operating system environments
  - High Performance component architecture
- **Unique Capabilities**
  - Large, coordinated research activities
  - Software testbed
  - Comprehensive character of research, with HPC focus
  - Strength of laboratory and academic participants
- **Ongoing Partnerships**
  - DARPA – HPCS planning and review team
  - DOD – Performance and productivity tools
  - NNSA – Open source support, reviews, planning
  - NSA – UPC, benchmark, programming model projects
  - HEC IWG/HECRTF – Interagency coordination leadership

Challenges in Data Management and Visualization

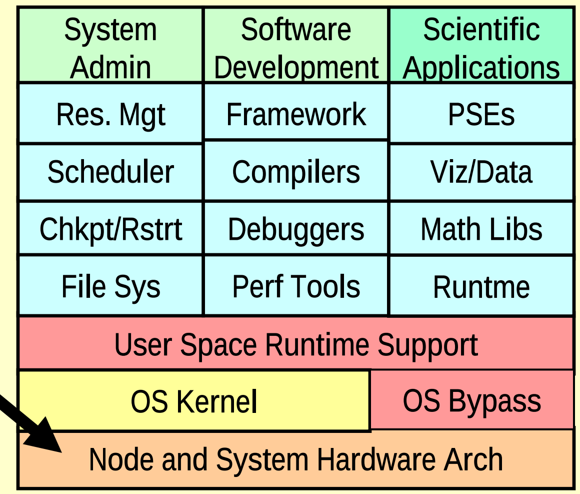
- R&D Goals:
  - Data representation, exploration, and understanding for terabyte to petabyte datasets
  - New ways to interact with data and experiments
  - Remote collaboration and remote visualization
- Outside program scope (mostly):
  - Scalable rendering
  - Infrastructure and support

Challenges in System Software

- **Petascale systems by 2008 (100,000 + processors)**
- **Challenging architecture diversity – X1, Red Storm, BG/L/P/Q, DARPA HPCS systems**
- **Reliability/fault management**

Petascale is different!

- **Too many nodes/processors for "brute force" approaches to debugging and performance analysis**
- **Multiple system failures will be normal**
- **Scalable software testbeds are essential**
- **Performance vs Productivity**



HPC System Elements

Meta challenge: Time from idea to robust, reliable software realization

How the Program Works – FASTOS

http://www.cs.unm.edu/~fastos/

• **February 2002:**	**(Wimps) Bodega Bay**
• **July 2002:**	**(Fast-OS) Chicago**
• **March 2003:**	**(SOS) Durango**
• **June 2003:**	**(HECRTF) DC**
• **June 2003:**	**(SCaLeS) DC**
• **July 2003:**	**(Fast-OS) DC**
• **March 2004:**	**Research Announcement**
• **2nd Half 2004:**	**Awards**
• **June 2005:**	**First annual PI mtg**
• **April 2006:**	**Special topic in ACM SIGOPS Operating Systems Review**
• **June 2006:**	**Second PI mtg at USENIX annual tech conf**

- 10 teams, about $21M invested over 3 years
- Part of HEC-URA, additional funding support from DARPA and NSA

FY2007 System Software Challenges and Plans

Drivers

- **2007 Budget Drivers**:
  - NERSC, +$17M, 100 – 150 TF NERSC-5
  - LCF@ANL, +$22M, 100TF IBM BG/P
  - LCF@ORNL, +$26M, 250 TF Cray XT4
- **Other drivers**
  - FY2007 – R&E Prototypes $13M
  - FY2008 – LCF@ORNL moves to 1PF

Challenge and Strategy

- **Challenge**
  - Leadership/production systems ready for full scale applications at acceptance
  - Fully capable/scalable/reliable vendor and DOE software
- **Strategy**
  - Documented plan, draft in progress
  - Application driven, selected by peer review
  - Focus on application-specific critical path software
  - Partnerships with vendors, labs and academia
  - Leverage NNSA experience and expertise
    - Sandia Cray Red Storm and Livermore IBM BG/L
  - Scalability testing
  - Comprehensive progress review
  - Adapt CS research program

DARPA HPCS Program and the DOE Office of Science

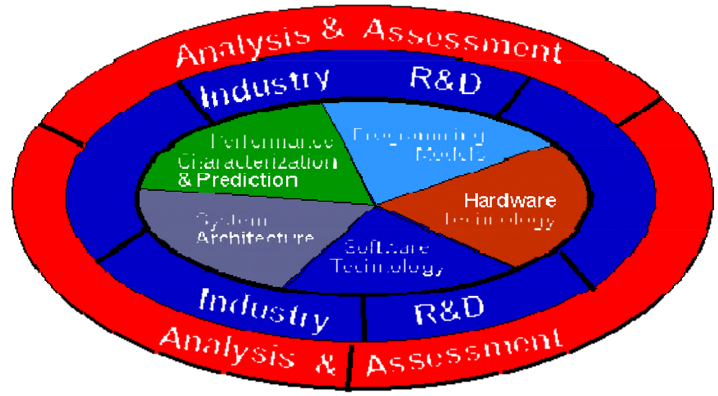
High Productivity Computing Systems

Goal:

- **Provide a new generation of economically viable high productivity computing systems for the national security and industrial user community (2010)**

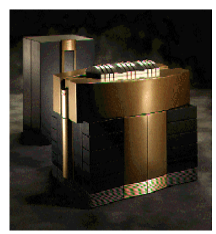
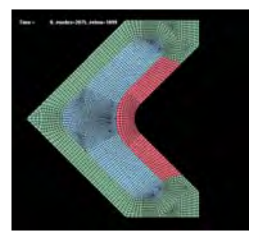
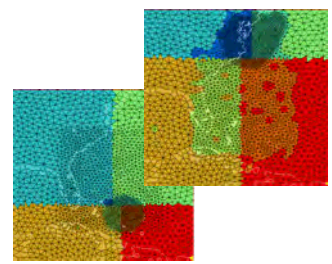
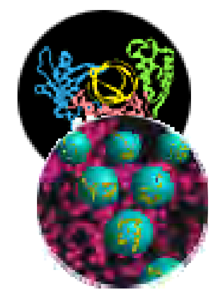
Impact:

- **Performance** (time-to-solution): speedup critical national security applications by a factor of 10X to 40X
- **Programmability** (idea-first-solution): reduce cost and time of developing application solutions
- **Portability** (transparency): insulate research and operational application software from system
- **Robustness** (reliability): apply all known techniques to **protect against outside attacks**, hardware faults, & programming errors



HPCS Program Focus Areas



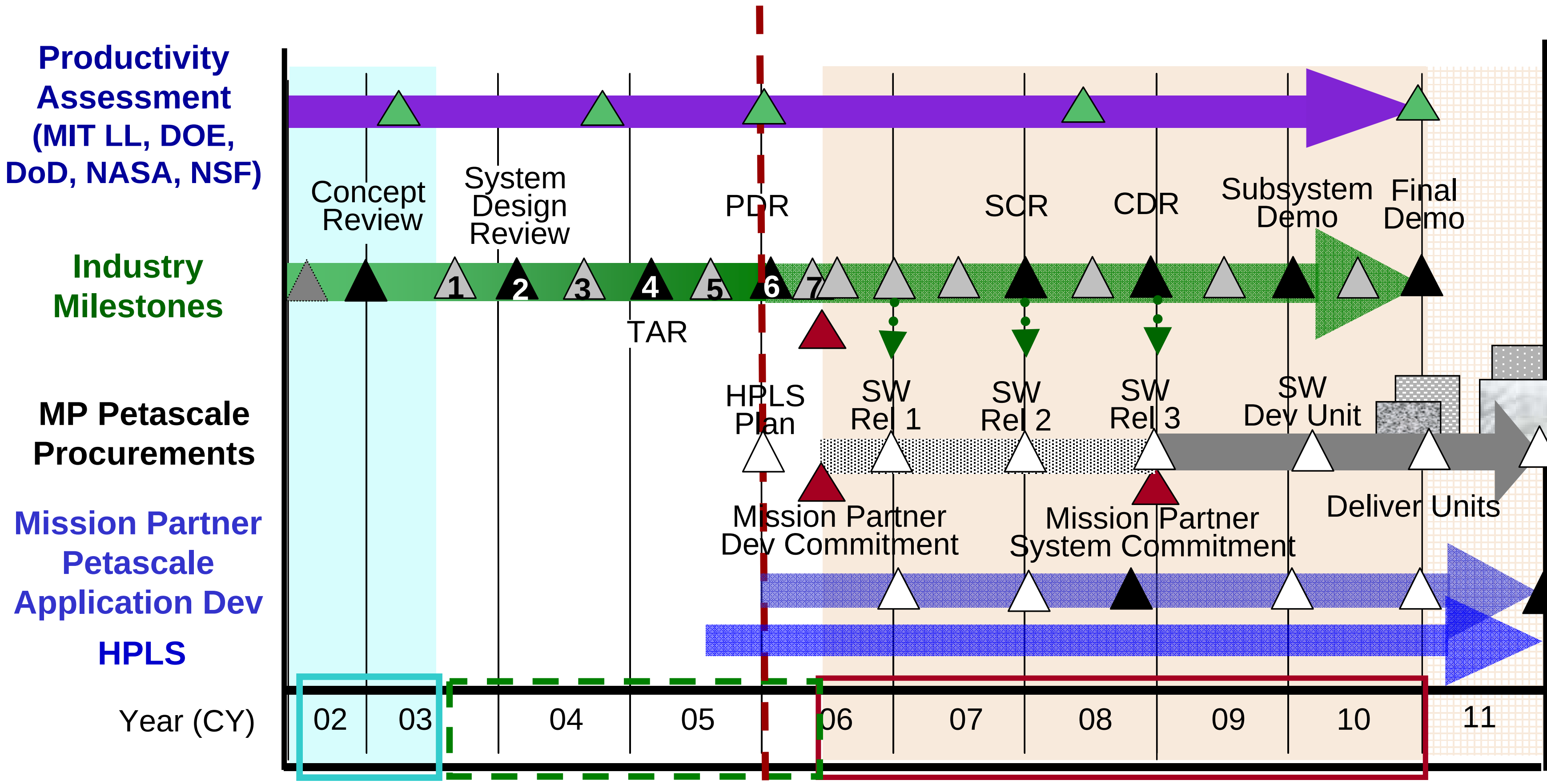
Applications:

- Intelligence/surveillance, reconnaissance, cryptanalysis, weapons analysis, airborne contaminant modeling and biotechnology

Fill the Critical Technology and Capability Gap

Today (late 80's HPC technology)…..*to*…..Future (Quantum/Bio Computing)

HPCS Program Phases I - III



Program Reviews

Critical Milestones

Program Procurements

(Funded Five)
Phase I
Industry Concept Study

(Funded Three)
Phase II
R&D

Phase III
Development and Prototype Demonstration

Mission Partners

Phase II Program Goals

- Phase II Overall Productivity Goals
  - **Execution (sustained performance) – 2 Petaflop/sec (scalable to greater than 4 Petaflop/sec). Reference: Workflow 3**
  - **Development – 10X over today's systems. Reference: Workflows 1,2,4,5**
- Productivity Framework
  - **Establish experimental baseline**
  - **Evaluate emerging vendor execution and development productivity concepts**
  - **Provide a solid reference for evaluation of vendor's Phase III designs**
  - **Early adoption or phase in of execution and development metrics by mission partners**
- Subsystem Performance Indicators (Vendor-Specified Goals)
  - **3.2 PB/sec bisection bandwidth;**
  - **64,000 GUPS;**
  - **6.5 PB/sec data streams bandwidth;**
  - **2+ PF/s Linpack**

10 to 10K times Delta from Business as Usual

Documented and Validated Through Simulations, Experiments, Prototypes, and Analysis

HPCS I/O Challenges

- **1 Trillion files in a single file system**
  - **32K file creates per second**
- **10K metadata operations per second**
  - **Needed for Checkpoint/Restart files**
- **Streaming I/O at 30 GB/sec full duplex**
  - **Needed for data capture**
- **Support for 30K nodes**
  - **Future file system need low latency communication**

An Envelope on HPCS Mission Partner Requirements

HPCS Status

- New Program Manager, Bill Harrod
- Phase III Solicitation release date: 3 March 2006
- Phase III Proposals due: 5 May 2006

ASCR and HPCS

- **Involved from Phase I, day 1**
  - Proposal review team
  - Progress review team
- **Phase II**
  - Proposal, Progress review teams
  - Mission partner
  - Support for Execution time productivity (Bob Lucas, USC/ISI
  - Support for Development time productivity (Jeremy Kepner, MIT/LL)
  - **For further info and HPC Challenge results:**
    - **http://www.highproductivity.org/**
  - Next Generation Programming models—Cray/Chapel, IBM/X10, SUN/Fortress (Rusty Lusk, ANL)
- **Phase III**
  - Four year $13M/year budget commitment
  - Focus as much as possible on system software ecosystem, details TBD
  - Petascale application development
- **Post HPCS**
  - Competitive procurement(s)