Working Group 7 Application-Driven System Requirements

Chair: Mike Norman

Vice Chair: John Van Rosendale

WG7 – Application-driven System Requirements Charter

- Charter
  - Identify major classes of applications likely to dominate HEC system usage by the end of the decade. Determine machine properties (floating point performance, memory, interconnect performance, I/O capability and mass storage capacity) needed to enable major progress in each of the classes of applications. Discuss the impact of system architecture on applications. Determine the software tools needed to enable application development and support for execution. Consider the user support attributes including ease of use required to enable effective use of HEC systems.
- Chair
  - Mike Norman, University of California at San Diego
- Vice-Chair
  - John Van Rosendale, DOE

WG7 – Application-driven System Requirements Guidelines and Questions

- Identify major classes of applications likely to dominate use of HEC systems in the coming decade, and determine the scale of resources needed to make important progress. For each class indicate the major hardware, software and algorithmic challenges.
- Determine the range of critical systems parameters needed to make major progress on the applications that have been identified. Indicate the extent to which system architecture effects productivity for these applications.
- Identify key user environment requirements, including code development and performance analysis tools, staff support, mass storage facilities, and networks.
- Example topics:
  - applications, algorithms, hardware and software requirements, user support

Discipline Coverage

- Lattice Gauge Theory
- Accelerator Physics
- Magnetic Fusion
- Chemistry and Environmental Cleanup
- Bio-molecules and Bio-Systems
- Materials Science and Nanoscience
- Astrophysics and Cosmology
- Earth Sciences
- Aviation

FINDING #1

It's all about us!

Top Challenges

- Achieving high sustained performance on complex applications becoming more and more difficult
- Building and maintaining complex applications
- Managing data tsunami (input and output)
- Integrating multi-scale space and time, multi-disciplinary simulations

Multi-Scale Simulation in Nanoscience

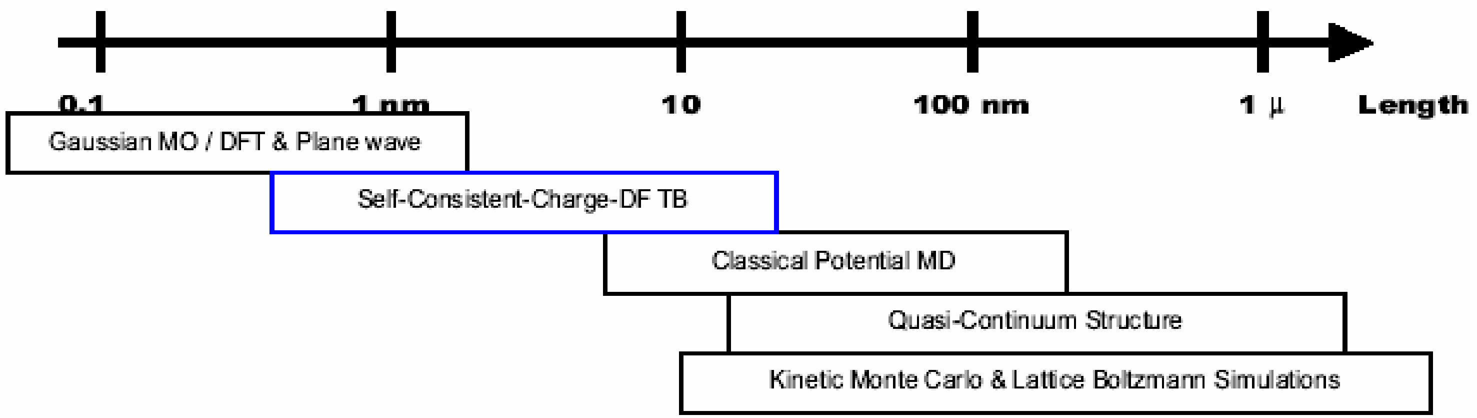
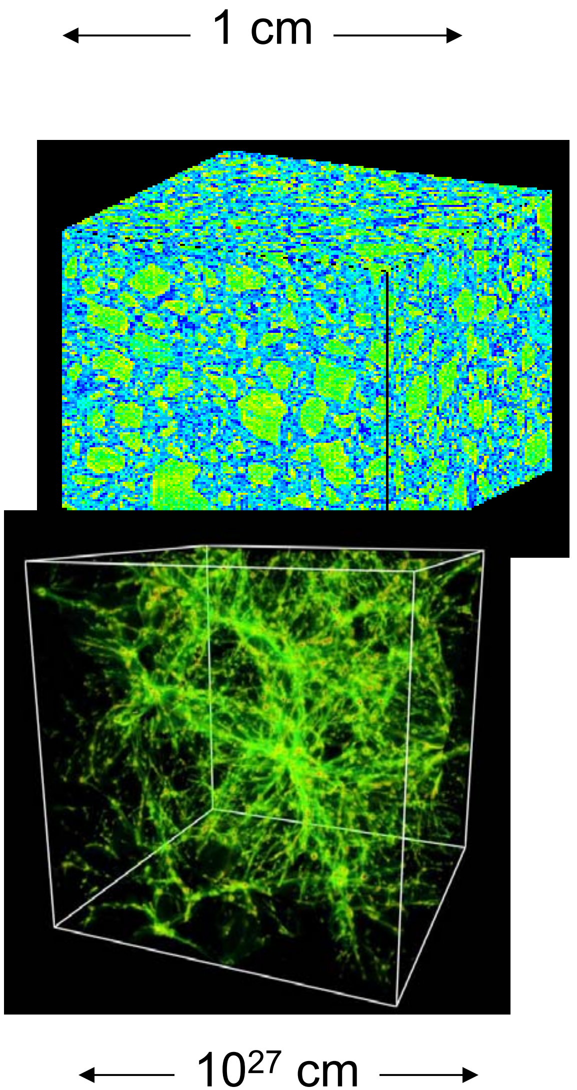


Fig. 1. Modeling And Simulations

Maciej Gutowski, WP 001

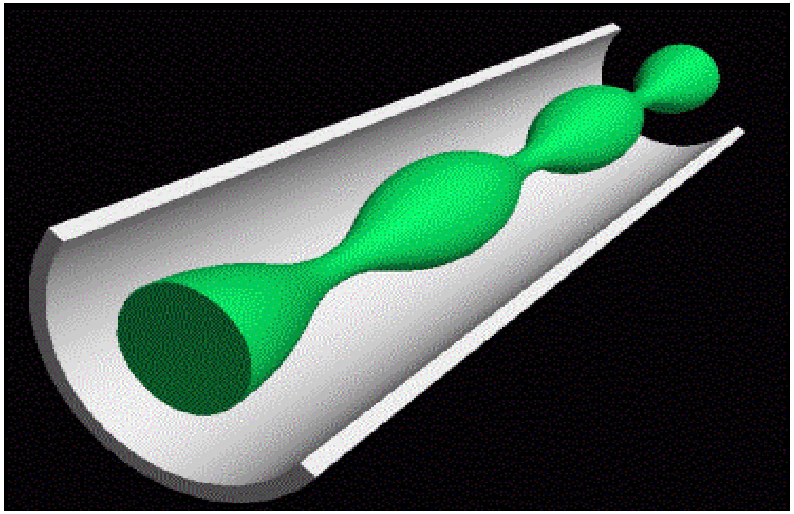
Question 1

- Identify major classes of applications likely to dominate use of HEC systems in the coming decade, and determine the scale of resources needed to make important progress. For each class indicate the major hardware, software and algorithmic challenges.



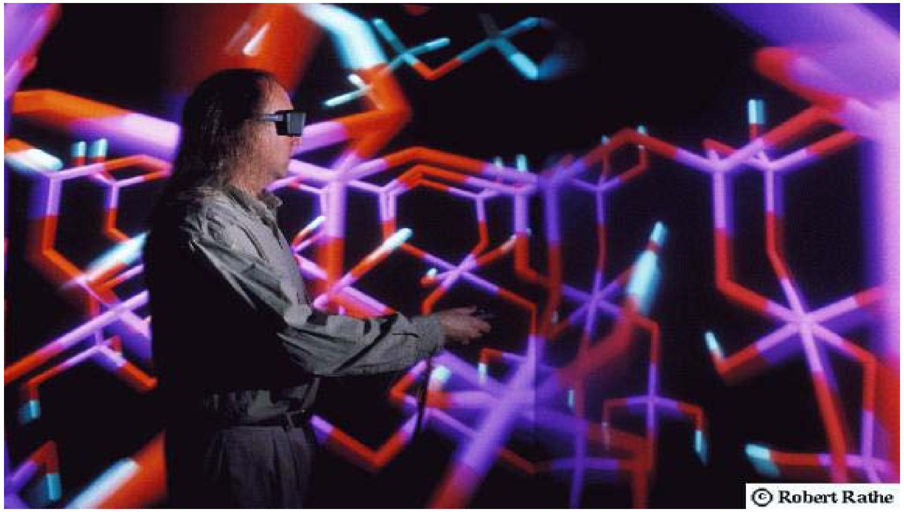
Question 2

- Determine the range of critical systems parameters needed to make major progress on the applications that have been identified. Indicate the extent to which system architecture effects productivity for these applications.



Question 3

- Identify key user environment requirements, including code development and performance analysis tools, staff support, mass storage facilities, and networks.



Findings: HW [1]

- 100x current sustained performance needed now in many disciplines to reach concrete objectives
- A spectrum of architectures is needed to meet varying application requirements
  - Customizable COTS an emerging reality
  - Closer coupling of application developers with computer designer needed
- The time dimension is *sequential*: difficult to parallelize – ultrafast processors and new algorithms are required.
  - fusion, climate simulation, biomolecular, astrophysics: multiscale problems in general

Findings: HW [2]

- Thousands of CPUs useful with present codes and algorithms; reservations about 10,000 (scalability and reliability)
  - Some applications can effectively exploit 1000s of cpus only by allowing problem size to grow (weak scaling)
- Memory bandwidth and latency seems to be a universal issue
- Communication fabric latency/bandwidth is a critical issue: applications vary greatly in their communications needs

Findings: Software

- SW model of single-programmer monolithic codes is running out of steam – need to switch to a team-based approach (a'la SciDAC)
  - scientists, application developers, applied mathematicians, computer scientists
  - modern SW practices for rapid response
- Multi-scale and/or multi-disciplinary integration is a social as well as a technical challenge
  - new team structures and new mechanisms to support collaboration are needed
  - intellectual effort is distributed, not centralized

Findings: User Environment

- Emerging data management challenge in all sciences; e.g., bio-sciences
- Massive shared memory architectures for data analysis/assimilation/mining
  - TB's / day (NCAR/GFDL, NERSC, DOE Genome to Life, HEP)
  - sequential ingest/analysis codes
  - I/O-centric architectures
- HEC Visualization environments a la DOE Data Corridors

Strategy and Policy [1]

- HEC has become *essential* to the advancement of many fields of science & engineering
- US scientific leadership in jeopardy without *increased and balanced* investment in HEC hardware and wetware (i.e., people)
- 100x increase of current sustained performance needed now to maintain scientific leadership

Strategy and Policy [2]

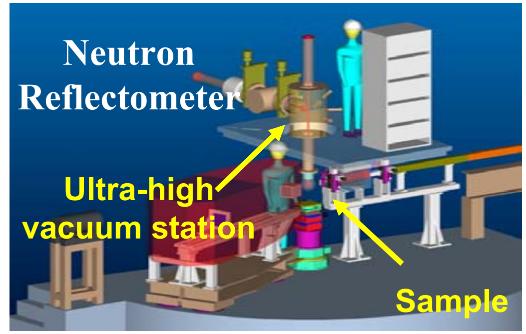
- A spectrum of architectures is needed to meet varying application requirements
- New institutional structures needed for disciplinary computational science teams (research facility model)
  - An integrated answer to Question 3

Facilities Analogy

National User Facility

User Interface "End Station"

Small Angle Scattering



High Res. Triple Axis

Users

- Polymer Science
- Nano-Magnetism
- Strongly Correlated Materials
- Dynamics

PSC

HPC Facilities

Fusion

Materials Science Research Network

Materials : Math : Computer Scientists

- *Standards Based - Tool Kits*
- *Open Source Repository*
- *Workshops*
- *Education*

QCD

Domain Specific Research Networks

- Fusion CRT 1
- Magnetism CRT
- Correlation CRT
- Microstructure CRT
- QCD CRT 1

Collaborative Research Teams

⇨ **Direction of competition**